

Anyone for neutrons?

There is a strong case for constructing a major new neutron source somewhere in Europe. The obstacles to progress towards this goal reflect much that is problematic about the organization of European science.

A negative hydrogen ion accelerates down a linear accelerator to 90% of the speed of light and punches through a carbon sieve, which strips it of its two electrons. The remaining proton enters a ring 70 metres in diameter, in which it and other protons are accumulated in pulses to be fired many times a second in a 5-megawatt beam onto 1,000 litres of mercury. Inside the mercury target, in a process known as spallation, the protons interact with the mercury's atomic nuclei, exciting them so that they 'evaporate', each releasing up to 30 fast neutrons. Some 30% of these neutrons escape through gaps in the shields that surround the mercury target, to be beamed towards one of dozens of possible destinations. These would incorporate bits of matter, scattered neutrons from which can be used to analyse its physical or magnetic structure, and its atomic or molecular dynamics, with a precision that is unavailable at any other facility anywhere.

This is the vision that underlies the proposed European Spallation Source (ESS). The ambition, or something like it, has been shared by scientists from many disciplines for ten years. Yes, they already have nuclear reactors and spallation sources to produce neutrons, but the limitations of these — in both power and lifetime — are becoming all too apparent. Yes, they have synchrotrons, but the ESS would complement future synchrotrons or free-electron lasers in its ability to probe particular regimes of size, energy and dynamics. Moreover, neutrons are powerful in distinguishing between elements and between isotopes, and can probe magnetic materials.

This week sees what could be called a landmark: representatives of major European countries are to sign a memorandum of understanding in support of the ESS. Neutron scatterers shouldn't hold their breath, however. The document effectively says no more than that the ESS, with two mercury targets (one for long pulses, one for short), is an excellent idea, and that more work will be done to develop an engineering design and a firm proposal to build it. One implicit, but so far unanswered, question is: where should one then send such a proposal?

Problems recognized

What is envisaged, and indeed required at these energies, is a facility to serve researchers on a continental scale. The United States, after abandoning plans for a national reactor source in 1996, finally decided in 1999 to embark on the construction of the Spallation Neutron Source at Oak Ridge, Tennessee — a single-target, 2-megawatt facility that is expected to be operational in 2006. Politicians, not least the then Vice President Al Gore, recognized the need for the United States to avoid a neutron drought and to recapture the cutting edge in neutron sources. Europe, with its reactor at the Institut Laue-Langevin in Grenoble and many smaller reactors, has traditionally had a healthier supply of neutrons to play with. Japan, too, has decided to push ahead with a new 1-megawatt spallation source, which should come online in 2007. All of this is in line with recommendations made in 1999 by the Organisation for Economic Co-operation and Development's Megascience Forum (as it was then known) to develop the next generation of spallation sources.

The new memorandum might be seen by optimists as a belated

recognition by European countries of the need for such a facility. But the reality is that the ESS remains bogged down in a quagmire of conflicting national interests, combined with a vacuum of structural leadership at the European level. In particular, the ESS's situation highlights the chronic inability of Europeans to provide credible infrastructure for themselves at a scale that requires multinational effort. The fact that it has taken ten years to reach the present stage of the ESS is a scandal in itself — and it is not for want of trying by the scientific enthusiasts behind the project (see www.ess-europe.de).

Inspection of the forces at play highlights the nature of the problem. The European Commission's research commissioner Philippe Busquin endorses the project as precisely the sort of facility that is needed for his European Research Area, and will help to prime it by supporting research and development for the proposal. But ultimately he has no line of expenditure that can be dedicated to the construction and management of such a facility. The European Science Foundation lacks the clout to be anything other than a catalyst, and has so far not even assumed this role.

Fine words

European nation states are where the funds will ultimately need to be found — but the current state of play does not look promising, despite their fine words of support. Germany is undertaking a lengthy review of all future big research projects under the auspices of its strategic body the Wissenschaftsrat, and gives no impression that the ESS will be championed by anyone with leverage. The French government is devoting much attention to the possibility of hosting the next fusion facility, ITER, whereas France's two major funding agencies, the CEA and the CNRS, have differing interests, leading to a lack of tangible support. The United Kingdom is under tight financial restrictions from the government that ensure that it could take years for anyone to stand up and pledge hard currency for the project.

Back in the mid-1990s, when faced with inertia that hindered the development of biotechnology in Germany, the research minister Jürgen Rüttgers introduced a competition between the Länder, triggering an explosion of biotechnology growth founded on regional self-interest. Cannily, the ESS project's leaders have pursued the same strategy, holding a competition for bidders seeking to host the facility.

Thus the development of a continental facility for neutron research — which will benefit universities, industries, physicists, chemists, earth scientists, materials researchers and biologists — now hangs on local bids. These have been made by a county in England (Yorkshire), two Länder in Germany (North Rhine-Westphalia and Saxony-Anhalt), a research centre in Germany (Jülich), a consortium of several Scandinavian centres, and the Rutherford Appleton Laboratory in Oxfordshire, UK. The fact that the fate of such a facility hangs on these local enthusiasms highlights the inadequacies of Europe's present research area. If the ESS is to be operating by its planned target of 2011 (already several years after the US facility is set to come online), national representatives in Europe — politicians, administrators and scientists — will need to get their act together fast. ■





Cell biologists
Married couple jailed
pending laboratory
theft charges
p886



Fast buck
CERN announces
cash-saving cutbacks
in synchrotron use
p888



Aiming high
New missions usher
in a revival for comet
exploration
p889



Film star
African predator
caught on camera
for the first time
p890

Code of conduct for bioethics branded 'soft' on corporate ties

Nell Boyce, Washington

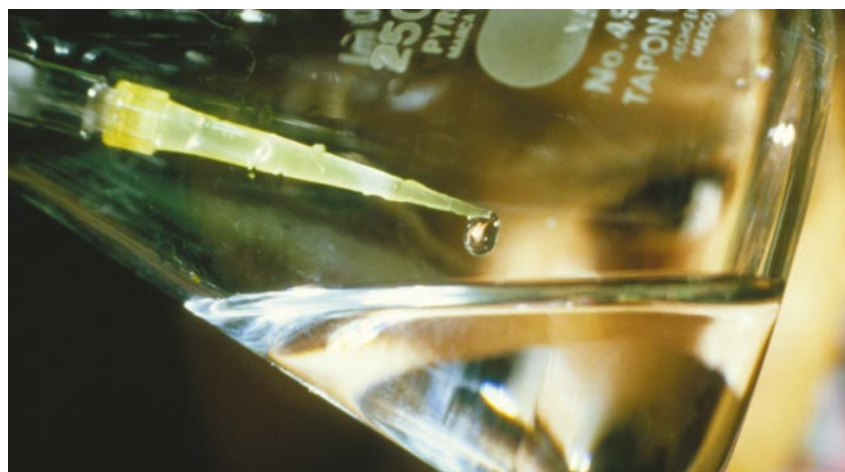
Big names in bioethics are falling out over how the burgeoning field should set its own ethical boundaries.

The controversy has been ignited by a report by a task force of ten leading bioethicists which outlines how bioethicists should interact with companies that seek their advice.

The task force was convened by two bioethics societies last year. Its report, in the May/June issue of the *Hastings Center Report*, suggests ways in which bioethicists can safeguard their independence when doing for-profit consulting. It covers issues such as disclosure, confidentiality contracts and compensation.

But in an accompanying commentary, bioethicists Stuart Youngner of Case Western Reserve University in Cleveland, Ohio, and Robert Arnold of the University of Pittsburgh in Pennsylvania dismiss the guidelines as "a 'how-to' manual for a new guild interested in promoting what it already does". They also say that task-force members have not disclosed their own corporate ties.

The dispute arises as bioethics is coming of age as a profession in the United States. As the field gains influence in the media and in policy-making, industrial companies have begun to engage bioethicists to assess controversial work — and to convince the public that they are taking ethics seriously.



Eyed with suspicion: some bioethicists are concerned by their peers' links with commercial science.

Bioethicists are being offered board positions, consulting contracts, research grants and even stock options. Some of them worry that, as a result, industry is co-opting the field. Others argue that corporate money poses no threat provided that bioethicists declare it and manage corporate relationships with care.

The *Hastings Center Report*, the journal of the bioethics centre of that name in Garrison, New York, began asking contributors to declare conflicts of interest last year. But the task force's own declaration merely says that

eight of the ten authors have "performed consultations of the type described in this article".

That's not good enough, say Youngner and Arnold. Their commentary argues that readers need more information to evaluate the panel's conclusions. To make their point, the two dissidents asked the journal to print exhaustive details of their own income from corporate sources — which it declined to do.

In an editorial, the journal says it doesn't think lengthy disclosures will benefit its readers, as "the facts would lose their punch" and

Bush's advisers lean towards ban on research cloning

Erika Check, Washington

The President's Council on Bioethics, which advises US President George W. Bush on thorny issues in medicine and biology, is edging towards recommending a moratorium on cloning human embryos for research.

At its meeting in Washington DC on 20 June, all of the council's members supported a ban on cloning to produce a human baby. But they were divided on 'therapeutic' cloning, which would create embryonic cells in order to study the early stages of human development. The majority of the council supports a moratorium of two to six years

on research cloning, but a few of its 18 members — including some scientists — have voiced opposition to such a move.

The council's chairman, philosopher Leon Kass of the University of Chicago, says that the panel's report, expected to be published this summer, will probably include a majority verdict and some dissenting views. "When he made his decision on stem cells, the president got as much information as he could," Kass says. "I don't think we'll do him a service if we don't make as strong a case as possible for the different sides" of the debate.

While the Senate has turned its attention

away from the cloning issue, Sam Brownback (Republican, Kansas), who leads opposition there to therapeutic cloning, has vowed not to let the issue die (see *Nature* 417, 775; 2002). Supporters of a moratorium say that backing from the bioethics council could help them garner the support they need.

Gene Tarne, a spokesman for a coalition of anti-cloning groups, says that most opponents of therapeutic cloning had not expected a firm recommendation from the council. If it backs a moratorium, he says, that could sway some senators to vote for it. ■

www.bioethics.gov



“authors would use the space to subtly spin those facts and forestall readers’ concerns”.

The co-chairman of the task force, Baruch Brody of Baylor College of Medicine in Houston, Texas, notes that science journals’ disclosure statements never include sums of money. He also doesn’t see the need for the task-force members to name companies with which they are associated in this case, because no firm stands to benefit from their published opinions.

“The purpose of disclosure is to put people on notice that there’s the potential for bias. We’ve done that,” says Brody. But he admits that confidentiality agreements stop him revealing some of his corporate relationships. He says he avoids potential conflict by never publicly discussing bioethics issues relating to those companies.

Such agreements worry Arnold because they rely on the bioethicist to assess where a conflict might arise. “I was saddened by the fact that the committee took a very soft position on secrecy agreements,” he says.

Arnold is also worried about problems posed by companies sponsoring academic centres. The Center for Bioethics at the University of Pennsylvania School of Medicine, directed by task-force member Arthur Caplan, for example, has taken funding from companies such as Celera Genomics of Rockville, Maryland, and Monsanto of St Louis, Missouri.



Arthur Caplan: declares funding.

But Virginia Sharpe of the Center for Science in the Public Interest, a pressure group based in Washington DC, says that Caplan’s is the only centre of 89 she surveyed that posts funding information on its website.

The bioethics societies have yet to act on the task force’s findings, which have split their community. “This has been the most divisive issue in the field as long as I’ve been in it,” says task-force member Jeffrey Kahn of the University of Minnesota.

♦ www.thehastingscenter.org

Scientists jailed for alleged theft from Harvard laboratory

Rex Dalton, San Diego

A married couple have been arrested and imprisoned in San Diego pending trial for their alleged theft of genetic material from Harvard University for shipment to an unnamed drug company in Japan.

Federal investigators say the couple stole genetic clones, reagents and records relating to immunosuppressive drug development from a Harvard molecular biology laboratory where they both worked in 1999.

Supporters of the pair in San Diego say that, although the couple made “stupid” mistakes at Harvard, they are being treated excessively harshly for doing what postdoctoral fellows often do.

The two researchers — Jiangyu “Jonathan” Zhu, a Chinese citizen, and his wife, Kayoko Kimbara, a citizen of Japan — or their attorneys could not be reached for comment on the case.

The couple, who were arrested on 19 June, are seeking release on bail, but the US Attorney’s office wants them to be detained and taken to Boston for trial.

After a two-year investigation by the Federal Bureau of Investigation (FBI), federal prosecutors issued a criminal complaint on 17 June, charging the pair with theft of trade secrets, interstate transportation of stolen property and conspiracy. These felony charges carry a maximum prison term of ten years and fines of up to \$250,000 each.

Zhu’s and Kimbara’s arrest is at least the third case of Asian biologists working in the United States being jailed pending theft charges in the past year (see *Nature* **411**, 225–226; 2001 and **417**, 576; 2002). But the cases were each handled by different law-enforcement agencies and so do not seem to be a part of an organized crackdown.

The couple met when they were postdocs in the Harvard Medical School laboratory of cell biologist Frank McKeon, who studies genetic control of the immune system. They allegedly stole materials and data relating to the genetic regulation of calcineurin, a signalling protein in the heart, brain and immune system.

An FBI affidavit claims that the pair switched to night shifts, worked without supervision and discovered potentially new immune-system genes, which were not disclosed to McKeon. After falling out with McKeon in 1999, the affidavit says that they moved in January 2000 to the University of Texas Health Science Center in San Antonio, where Zhu was made an assistant professor with Kimbara as his postdoc.

The pair removed more than 30 boxes of



Jiangyu Zhu (left) faces up to ten years in prison for allegedly stealing from Frank McKeon’s Harvard lab.



materials from McKeon’s lab late at night, at the end of 1999, according to the affidavit. In early 2000, it claims, Harvard officials went

to San Antonio, where they searched Zhu’s lab, locating and repossessing Harvard material that they valued at \$300,000. Texas officials then declined to renew Zhu’s contract, prompting him and his wife to leave that summer.

Unable to secure a research position, Zhu contacted Jean Wang, a molecular biologist at the University of California, San Diego (UCSD), who says she decided to give him “a second chance” in January 2001.

While Zhu began a new line of research on apoptosis in thyroid cancer cells in Wang’s laboratory, Kimbara took up a postdoctoral position at the Scripps Research Institute in La Jolla, working in the vascular-biology lab of Mark Ginsberg.

Wang, who consulted McKeon before hiring Zhu, feels that a double standard is being applied to the couple, as postdocs often take reagents and records when they leave a laboratory. “It is horrible,” says Wang of the prosecution. “It is like a goliath crushing an ant.”

Wang acknowledges, however, that she wasn’t aware of the alleged transfer of research material to Japan until this May. After she learned of the FBI inquiry into a transfer, Wang says Zhu told her he sent plasmids to a former McKeon postdoc who had returned to his native Japan. This was described as an innocent scientific exchange to create antibodies, Wang recalls.

But federal records allege that e-mails show that the material was destined for commercial use for a possible drug to block organ rejection. The Japanese company — which has not been named — is cooperating with the US authorities and has returned the Harvard material.

AP/UCSD

Fight against terror 'needs science institute'

Erika Check, Washington

The US National Academy of Sciences has called on the government to create a new institute for homeland security to assist policy-makers in the fight against terrorism.

The academy also asks for the creation of a new granting arm of the National Institutes of Health to fund high-risk research, along the lines pioneered by the Defense Advanced Research Projects Agency (DARPA).

The requests are made in a report that the academy, together with the National Academy of Engineering and the Institute of Medicine, has been painstakingly preparing since shortly after the terrorist attacks of 11 September last year. The report, which was released on 24 June, also recommends a wide range of scientific measures to defend the country from nuclear, biological, chemical, computer and infrastructure attacks.

The institute would, the report says, function as a contractor to the government, and would help to test and set standards for devices to prevent and contain terrorist attacks. These include sensors that would help emergency workers, food and shipping inspectors and city officials to thwart biological, chemical and nuclear attacks.

"There is going to be a proliferation of technologies and it will be important to characterize how they work and how they can best be deployed," says Richard Klausner, a former director of the National Cancer Institute and co-chair of the panel that wrote the report, who now works for the Bill and Melinda Gates Foundation in Seattle, Washington.

As well as the new granting mechanism at the NIH, the report suggests that the Department of Agriculture should support a new funding programme to coordinate the

nation's response to plant pathogens. It also calls for the United States to take extra steps to recruit and train the next generation of scientists and engineers.

The report adds that the government should do more to track nuclear weapons and other radioactive materials, and to provide more funding to help Russia dilute its supply of highly enriched uranium. And it calls for more research in many areas, including the psychological response to terrorism, the human immune system, and the ability of contaminants to permeate the water supply.

The report does not suggest a budget for any of these initiatives. And as it was not requested by the government, some have questioned its likely impact. But Congress seems to be receptive to the report, and committees in both the House and the Senate have arranged hearings on its findings.

Some researchers complained that the document was silent on issues such as how to deal with the publication of what it calls "sensitive but unclassified" research. The report says that the White House Office of Science and Technology Policy should begin dialogue with scientists on this issue, but does not recommend any particular approach to it.

Critics also argue that the report misses the opportunity to urge better integration of science into US intelligence agencies, which have been criticized for their failure to prevent last September's attacks.

Much of the report was drafted before 6 June, when President George W. Bush introduced plans to create a new Department of Homeland Security. But researchers hope that it will help to set the scientific agenda for the department, which Bush promised would coordinate research against terrorist threats. ■



Prominent US scientists have called for research to help prevent a repeat of 11 September's attacks.

Lack of funds puts 'bubble fusion' replication on hold

Geoff Brumfiel, Washington

Efforts to replicate an experiment that detected 'bubble fusion' in a small jar are struggling to win funds, prolonging a controversy over the validity of the result.

When the original experiment was published in March (*Science* 295, 1868–1873; 2002), it sparked a keen debate over whether nuclear fusion had actually occurred. A team led by Rusi Taleyarkhan, a nuclear engineer at Oak Ridge National Laboratory in Tennessee, claimed success based on measurements of two major by-products of fusion, neutrons and tritium.

But critics said that an internal review at Oak Ridge six months earlier had called those measurements into question. Recalling the discredited 'cold fusion' experiments a

decade earlier, researchers asked for an independent effort to replicate the results.

Lee Reidinger, deputy director of Oak Ridge, convened a multidisciplinary team, including Taleyarkhan, to refine the original set-up and use better detectors. The effort was expected to cost up to \$500,000.

Reidinger says he hoped to use part of Oak Ridge's budget that is reserved for research conducted at the laboratory's own discretion. But he says that officials at the Department of Energy told him this could only be used for original experiments, not for verifying existing work. Now the group is to submit a formal proposal for funding — a process that will take several more months, with uncertain prospects of success.

Taleyarkhan, meanwhile, has won a new

grant from the Defense Advanced Research Projects Agency (DARPA), believed to be worth around \$250,000, to continue his own investigation. The funding from DARPA, which specializes in high-risk projects, will be used to scale up his apparatus. Taleyarkhan will be looking at the process's potential not just for fusion power, but also for "more immediate applications" as a neutron source, perhaps for radiography or detection of explosives.

He also says he is helping several groups worldwide to reproduce his original result, although he declined to identify them for fear that publicity could adversely affect their funding proposals. "I believe it will take at least four to six months before you start seeing results," he adds. ■

CERN puts research on hold to build collider

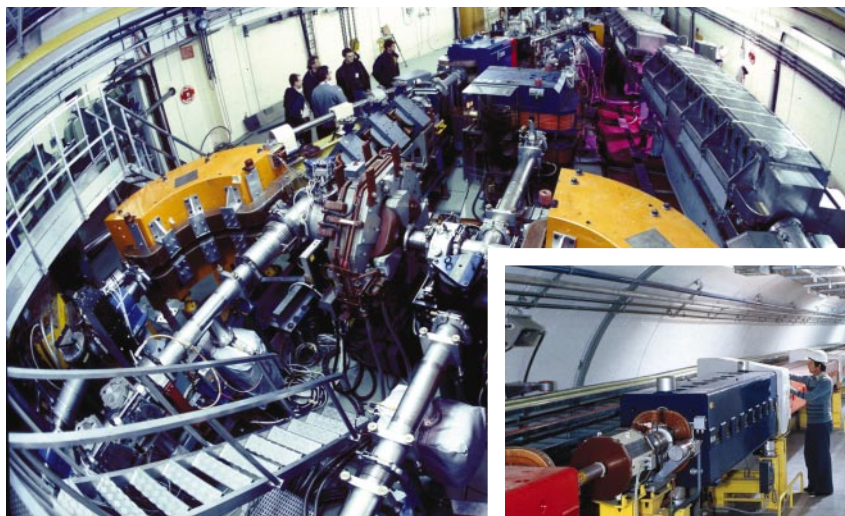
Alison Abbott, Munich

CERN, the European particle-physics laboratory in Geneva, Switzerland, is to shelve most of its medium-term research plans in a bid to ensure the completion of its main project — the construction of the Large Hadron Collider (LHC).

Under a retrenchment plan agreed on 21 June between the laboratory's management and its governing council, physicists at CERN will generate practically no fresh experimental data during 2005. Some researchers have already expressed fears about the impact of this data drought on the laboratory.

Cost overruns totalling SFr850 million (US\$570 million) on the LHC's SFr2.6-billion construction cost were exposed last autumn (see *Nature* **413**, 557; 2001). Council members, whose governments foot CERN's bill, were angry that its director, Luciano Maiani, had known about the problem for months but had not informed them.

In September, Maiani set up internal task forces to consider improvements in CERN's management and to find ways of saving money to pay for completion of the LHC. CERN's council, meanwhile, established an external review committee, chaired by Robert Aymar, director of ITER, the inter-



Cost cutting: the Proton Synchrotron (above) and Super Proton Synchrotron will be shut down for all of 2005.

national project to build an experimental magnetic fusion reactor.

Both processes reached similar conclusions, and CERN's council has now accepted Maiani's plan to implement these by streamlining the laboratory's management, winding

down small research projects and adjusting the LHC construction plan.

The LHC, which physicists hope will find the Higgs boson, will not now come on line until 2007, two years later than originally planned. Loans have been arranged to extend the period of payment for the LHC until 2009.

Work not related to the LHC is being cut back to concentrate CERN's resources on the collider. Running time at the Proton Synchrotron and the Super Proton Synchrotron, which provide proton beams for a range of experiments, will be cut by 30% until 2004, and the machines will be switched off for the whole of 2005, saving SFr34 million.

"The loss of beam time and resources will cause problems for high-energy physicists in Europe and beyond, but we will just have to suffer for a limited number of years," says Maurice Bourquin, a particle physicist who is rector of the University of Geneva and also president of CERN's council.

"Of course it is always painful when data stop flowing," says Maiani. "But it is better to make hard decisions than to cultivate illusions." He adds that there are compensations for scientists not connected with the LHC. The accelerators will be refurbished during the close-down, and the resumption of programmes in 2006 will be guaranteed explicitly in CERN's budget.

But some scientists at CERN are not satisfied. Stephan Paul, a spokesman for the Compass programme that studies the structure and function of hadrons, says that "diversity is required to keep a rounded view of high-energy physics". Until now, diversity was a key feature of CERN's culture. "It's a tragedy that it's being eroded," says Chris Tully, a physicist at Princeton University who works on one of the LHC detectors.

Ocean carbon study to quit Hawaii

Virginia Gewin, Washington

An international consortium is withdrawing its application for a permit to conduct a large carbon-sequestration experiment off the coast of Hawaii, in the face of opposition from environmentalists.

The researchers, who plan to release tonnes of liquefied carbon dioxide in the deep ocean, say there are no environmental grounds for dropping the \$5 million project. They plan to decamp to Norway, and expect to find out in the next few weeks whether they have permission to do the work there.

Results from the experiment are expected to shed light on the feasibility of carbon sequestration in the oceans as a means of mitigating the build-up of carbon dioxide in the atmosphere, which most researchers think causes global warming.

The time taken to secure a permit "was becoming excessive", says Gerard Nihous, an ocean engineer at the Pacific International Center for High Technology Research in Honolulu, which was due to carry out the study. He says the consortium plans to notify the US Environmental Protection Agency (EPA) this week that it is withdrawing its permit application.

The amount of carbon dioxide to be

released in Norway will be less than the 60 tonnes envisaged for release in Hawaii, but it will still be the largest experiment of its type, the researchers say.

The project, which is supported by Japan, the United States and Norway, was conceived in the 1997 meetings that led to the Kyoto Protocol. But local environmental groups claimed that the experiment would acidify Hawaii's fishing grounds (see *Nature* **401**, 315; 1999).

Environmentalists are suspicious of the experiment because they see carbon sequestration as a 'red herring' introduced into the Kyoto negotiations to divert attention from the need to reduce greenhouse-gas emissions. "We're against the study because we think it's part of a larger effort that is looking for a panacea that will encourage more fossil-fuel use," says Jeffrey Mikulina, director of the Hawaiian chapter of the Sierra Club, a large US environmental group.

But an assessment performed for the EPA said that the experiment would have no substantial environmental impact. "If we halt small experiments like this, we're never going to get the answers," says Howard Herzog of the Massachusetts Institute of Technology, one of the study's investigators.

NASA launch heralds fresh wave of comet exploration

Tony Reichhardt, Washington

The planned launch of NASA's CONTOUR (Comet Nucleus Tour) spacecraft in Florida on 1 July will herald a period of intense comet exploration, the like of which has not been since an armada of spacecraft visited Halley's Comet in 1986.

Between now and 2006, four spacecraft will either set off for, or return data from, the vicinity of a comet, capturing close-up images as well as dust and gas samples.

Three NASA missions — CONTOUR, Stardust and Deep Impact — as well as the Rosetta mission, launched by the European Space Agency (ESA), will "take the investigation of comets to a completely different stage", says Hal Weaver of Johns Hopkins University in Baltimore, Maryland. Spacecraft cameras and spectrometers are far better than they were 16 years ago, points out CONTOUR's principal investigator, Joseph Veverka of Cornell University in Ithaca, New York, and are better suited to answering questions about the chemistry and formation of comets.

CONTOUR will pass through the comas — clouds of dust and gas — that surround the icy nuclei of the Encke and Schwassmann-Wachmann 3 (S-W3) comets in 2003 and 2006, respectively.

The craft's mass spectrometer will measure the ratio of hydrogen to deuterium in comets, revealing whether they delivered water to Earth early in the planet's history.

Ground-based studies of two comets that recently passed close to Earth, Hale-Bopp and Hyakutake, suggest that they may not have. But both of these originated in the distant Oort Cloud that surrounds the Solar System, and may not have the same chemical make-up as Encke, S-W3 and other short-period comets, which came from the much nearer Kuiper belt outside Pluto's orbit.

Only two cometary nuclei have been photographed at close range — Halley's Comet, by ESA's Giotto spacecraft in 1986, and Borrelly, last year by NASA's Deep Space 1. Neither was able to show conclusively whether the comets' nuclei were coherent or loose "rubble piles", says Weaver. But CONTOUR's images will have 10 times the resolution of Deep Space 1's, and 25 times that of Giotto's.

NASA's Deep Impact probe, which is scheduled to launch in 2004, will go a step further, smashing into the comet Tempel 1 a year later to study the material kicked up by the impact. Stardust, which has already been launched, will bring cometary dust back from the comet P/Wild 2 in 2006.

But the most expensive and ambitious of the new cometary explorers is ESA's US\$700-million Rosetta, which was a partnership with NASA before the US agency withdrew in 1996 (see *Nature* **383**, 469; 1996). Rosetta will be launched next January and reach the comet Wirtanen in 2011, where it will place one probe in orbit and land another on the nucleus. The mission aims to observe the comet as it heats up on its approach to the Sun and begins to spew gas and dust.

If Rosetta is the most thorough investigation on the drawing board, CONTOUR may be the most flexible. It was born, says Veverka, out of US scientists' frustration at not mounting a large comet mission in the 1980s and 1990s, despite several aborted NASA plans. By design, CONTOUR's orbit will bring the craft looping past the Earth once a year.

When it finishes its primary mission, it can easily be targeted to meet with other known comets, including d'Arrest, which is younger than Encke but older than S-W3. The ideal scenario would be for a new comet to be thrown towards Earth from the Oort Cloud, giving scientists the chance to study a different class of object at close quarters.

Any encounters after those with Encke and S-W3 would need to be approved by NASA. CONTOUR's total project cost was US\$159 million, half the price of a typical mission in NASA's Discovery series of planetary probes. So comet specialists are hoping that the agency will be generous if they ask it to prolong CONTOUR's life beyond 2006. ■



Thumbs up: Claudie Haigneré's appointment is broadly welcomed by French researchers.

Astronaut lands top research post in new French government

Sally Goodman, Paris

The astronaut Claudie Haigneré has been named minister for research in France's new centre-right government.

Haigneré's scientific background includes a medical degree and a PhD in neuroscience. Haigneré the astronaut is a familiar face to the French public, having first gone into space in 1996 to the Mir space station. Last October she spent ten days on the International Space Station.

Her high profile has helped to persuade researchers that she is suitable for the job despite her lack of political experience. "It's an original and courageous choice of minister," says Pierre Tambourin, director of the Genopole science park in Evry, near Paris. "I think she will defend science well."

Haigneré will have clear authority to tackle the most pressing problems facing French research. Jean-Pierre Raffarin, the new prime minister, belongs to the same party as president Jacques Chirac, ending the power-sharing between president and parliament that previously stifled reforms.

Improving the attractiveness of scientific careers and increasing the competitiveness of French research are likely to be high on her list of challenges.

Chirac has promised that research will be a priority. "We hope he sticks to his word," says Tambourin. "Researchers have not forgotten previous occasions when Chirac was in power when he reduced the research budget to fund other priorities."

Haigneré's research responsibilities will be overseen by Luc Ferry, head of the new superministry of youth, education and research (see *Nature* **417**, 214; 2002). But Ferry will be in charge of higher education — even though most labs run by the CNRS, France's basic-research agency, are hosted by universities. ■

NASA



Exploring new frontiers: the CONTOUR craft will pass close to the Encke and S-W3 comets.

Singapore joins band of countries cloning embryos for research

Tokyo Singapore looks set to join the short list of nations that sanction the creation of human embryos for research. Therapeutic cloning will become legal in the island state if, as expected, the government accepts the recommendations of its bioethics committee, which were released on 21 June.

The committee, which was chaired by Lim Pin, the former vice-chancellor of the National University of Singapore who is now a consultant at the National University Hospital, also said that a new regulatory framework was needed to ensure informed consent when collecting human embryos, eggs or cells, that the sale of donated materials should be banned, and that tissue from dead fetuses could be used in research.

The new policy could play a part in Singapore's plans to establish itself as a leading centre for life-sciences research (see *Nature* **412**, 370–371; 2001). The government is investing heavily in developmental biology and biomedical research, and has recently attracted scientists such as Alan Colman, a member of the team that created Dolly the sheep, the

first cloned mammal, at the Roslin Institute in Edinburgh, UK.

Mexico's modified maize under environmental study

Montreal The economic and environmental implications of the growth of genetically modified maize in Mexico are to be studied by a North American environmental agency.

The transgenic maize near Oaxaca in Mexico created an uproar when it was revealed in September (see *Nature* **413**, 337; 2001). The commercial growth of genetically modified corn is banned under a 1998 government moratorium, and environmental groups and scientists say that its presence could threaten the biodiversity of maize in the area.

The North American Commission for Environmental Cooperation, which studies environmental issues relating to trade between Mexico, the United States and Canada, announced on 20 June that it is to study the implications of the contamination. Its report will be issued early next year.

♦ www.cec.org

Shy predator comes out of the shadows

Dar es Salaam The first ever photograph of the rare Lowe's servaline genet, a mammal

related to the mongoose, has been snapped by ecologists working in Tanzania.

The genet, which is extremely reclusive, is thought to live in trees and forage at night. However, ecologists know little else about the three-foot-long predator. It has not been seen for 70 years, and much knowledge of it comes from one pelt that was collected in 1932.

But earlier this month, Daniela De Luca of the Wildlife Conservation Society, based at the Bronx Zoo in New York, caught the animal on film while studying carnivores in the Udzungwa Mountains National Park in central Tanzania.

The genet is named after its discoverer, the British explorer Willoughby Lowe. Another of Lowe's discoveries, the Miss Waldron's red colobus monkey, was declared extinct in 2000. De Luca hopes to save the genet from a similar fate, and plans to conduct ecological



Film premiere: the reclusive Lowe's servaline genet is captured on camera for the first time.

D. DE LUCA/WILDLIFE CONSERVATION SOCIETY

studies and interviews with local people to find out more about its habits.

Windy cities cause a storm for rain-soaked residents

Washington A NASA satellite has revealed that areas downwind of large cities receive almost a third more rain than those upwind.

This 'heat island effect' first emerged from rain-gauge recordings around St Louis, Missouri, in the 1970s. Buildings and roads retain more heat than trees and lawns, and the rising warm air is thought to push moisture to higher altitudes, where it condenses into clouds. These move downwind and empty their contents onto unfortunate suburbanites.

To try and confirm the effect, meteorologists at NASA's Goddard Space Flight Center in Greenbelt, Maryland, studied data from the Tropical Rainfall Measuring Mission, which launched in 1997. With a resolution of only 25 kilometres, the satellite's on-board radar is a relatively crude tool for spotting such localized rainfall differences, says project leader Marshall Shepherd. Even so, 28% more rainfall on average was recorded downwind for all five cities studied, which were in Georgia, Alabama and Texas. (J. M. Shepherd, H. Pierce & A. J. Negri, *J. Appl. Meteorol.* 41, 689–701; 2002).

Positive opening for negative results

Washington A null result may no longer mean no publication, at least for biomedical researchers. *The Journal of Negative Results in Biomedicine* is to be launched online later this summer by BioMed Central, a London-based company that publishes free-to-access life-science journals. Edited by Bjorn Olsen, a cell biologist at Harvard Medical School, the journal will be dedicated to publishing basic science and clinical-trial results on unsuccessful drugs or treatments.

Negative results often contain valuable information, but fail to find their way into electronic or print publications. Publishing failed work may not set researchers' pulses racing, but the publishers say that making the information easily accessible will help to prevent duplicated efforts.

Ancient rust-proofing shows Indians' metallic mastery

New Delhi The mystery of how a 1,600-year-old wrought-iron pillar has survived the high humidity of Delhi without rusting has been solved by a metallurgist from the Indian Institute of Technology. The pillar was erected by Kumara Gupta I of the Gupta dynasty that ruled northern India in

AD 320–540. It is over seven metres in height and weighs more than six tonnes.

R. Balasubramaniam, from the institute's branch in Kanpur, has now shown that the lack of rust is due to a high level of phosphorus in the iron, which catalyses the production of a protective layer of misawite, a compound of iron, oxygen and hydrogen (see R. Balasubramaniam *Curr. Sci.* 82, 1357–1365; 2002). He believes that the layer, which is just a few hundredths of a millimetre thick, began to form within three years of the pillar's erection.

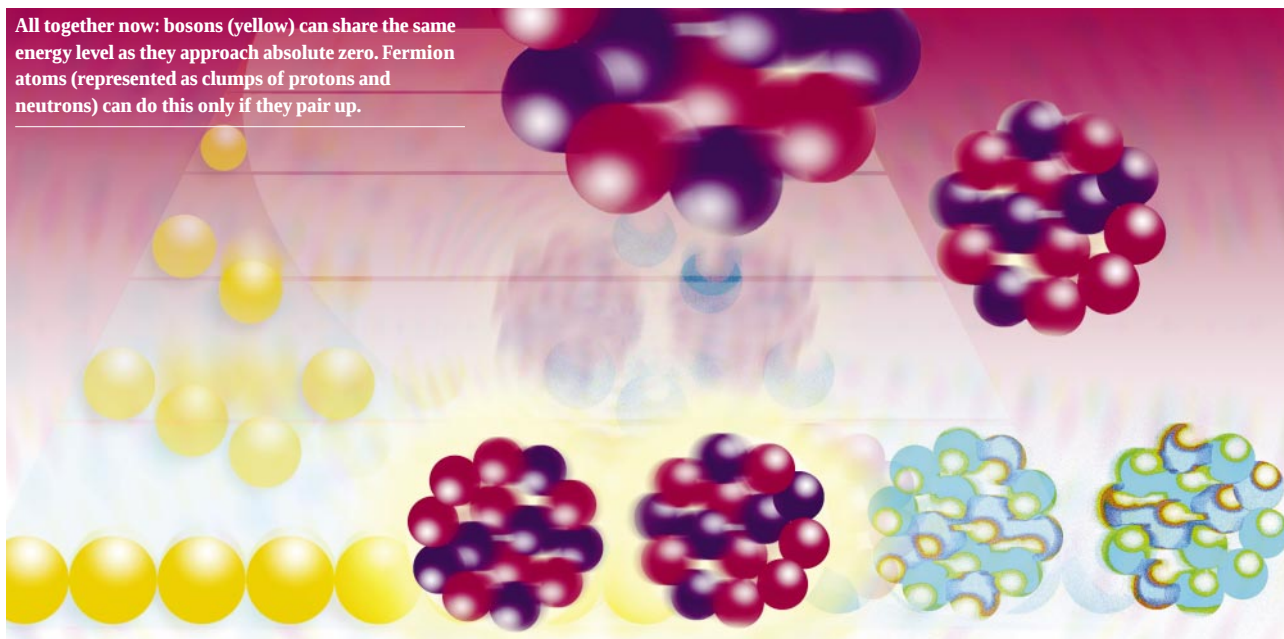
"The pillar is living testimony to the skill of metallurgists of ancient India," says Balasubramaniam.



Pillar of the community: cunning chemistry has preserved this Indian metal pole for centuries.

D. CUMMING/EYE UBIQUITOUS/CORBIS

All together now: bosons (yellow) can share the same energy level as they approach absolute zero. Fermion atoms (represented as clumps of protons and neutrons) can do this only if they pair up.



M. XERIDAT

Altered states

During the 1990s, ultracold gases were used to open up a new and often bizarre frontier of physics. Now researchers are poised to use similar gases to enter another, equally intriguing, realm. Mark Haw reports.

Stand by for a new state of matter — according to some predictions, it could be created in the next few months. In a handful of laboratories around the world, physicists are trying to persuade antisocial atoms to pair up to form a new type of ‘superatom’.

If they succeed, their creations will have more than curiosity value. The state — called a fermion condensate — would be perfect for testing theories of quantum mechanics, and could even lead to new ways of computing. Those involved in the effort know they are onto something big. “This state is unlike any other,” says Murray Holland, a theoretical physicist at the University of Colorado at Boulder, “and we have a physical system in which it may be realized.”

The key is to cool gaseous atoms to very low temperatures. An atom’s energy is limited to a series of discrete values called energy levels. As a gas reaches very low temperatures, the atoms slot themselves into the few remaining levels above absolute zero.

Different atoms do this in different ways. Those that belong to the boson family of particles, such as atoms whose total number of electrons, neutrons and protons is even, are perfectly comfortable occupying the same energy states as their neighbours. When a gas of bosons is cooled to close to

absolute zero, the particles all fall into the lowest energy state.

In quantum mechanics, particles can be represented as waves, and when all the atoms fall into a single energy state, their ‘wavefunctions’ fit together so that the peaks and troughs match up. The entire gas behaves as a single superatom called a Bose–Einstein condensate (BEC). Such a system was first created in 1995 by Carl Wieman, Eric Cornell and their colleagues at the University of Colorado at Boulder¹.

Antisocial attitudes

Fermions, the family of particles that includes atoms whose total number of particles is odd, are less friendly. Quantum mechanics says that no two fermions can share the same quantum state. As gases of fermions are cooled, the atoms play a game of musical chairs as they compete to secure their own energy level.

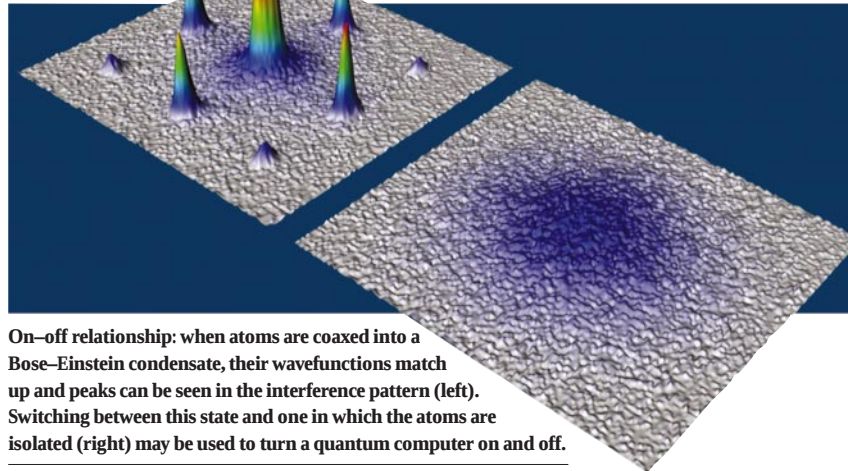
This makes creating a fermion version of a BEC sound impossible. But since the 1950s, physicists have known that, under the right conditions, two fermions can pair up. Take electrons — which are also fermions — in a superconductor. The vibrations of the lattice of atoms through which the electrons move constrains their motion in such a way that the wavefunctions of the individual

electrons pair up, and correlations form between them. In terms of quantum mechanics, the electrons form a kind of composite particle, even if they are not spatially close together. At low temperatures, this effect is strong enough for all electrical resistance, which is caused by unpaired electrons, to disappear.

In this case, the electron pairs behave as a single composite boson. So physicists reason that if pairs of fermionic atoms could be paired up in a similar way, they too should act like bosons and, in theory, form fermionic versions of BECs.

This is a tantalizing possibility. BECs have so far been used to test the details of quantum mechanics², to model exploding stars³ and to bring light to a halt⁴. Fermion condensates should open up a whole new range of experiments. And to get the atoms to pair up, physicists are learning how to take direct control of atomic interactions. By doing so, they could find new ways to manipulate matter and explore quantum phenomena.

But getting fermions to pair up in a gas, where there is no atomic lattice to constrain their motion, is a big challenge. The atoms occasionally pair up when they collide. Lowering the temperature, which reduces the speed of the atoms, increases the likelihood



On-off relationship: when atoms are coaxied into a Bose-Einstein condensate, their wavefunctions match up and peaks can be seen in the interference pattern (left). Switching between this state and one in which the atoms are isolated (right) may be used to turn a quantum computer on and off.

of pairs forming. But even the ingenious methods used to create BECs, in which the atoms of a gas are cooled by using lasers to slow them down, cannot create temperatures low enough for large numbers of fermion pairs to form.

So physicists plan to give the fermions a little push. The favoured method is to use a magnetic field to tweak the energy levels of the electrons orbiting the atoms. Adjusting the levels in the right way increases the chances that collisions will result in pairings.

Take your partners

Last month, Neil Claussen and his colleagues at the University of Colorado used this trick to pair up bosons⁵, and the team is now trying to do the same with fermions. Groups led by Wolfgang Ketterle at the Massachusetts Institute of Technology and Randall Hulet at Rice University in Texas are taking similar approaches.

Together with Henk Stoof, a theorist at Utrecht University in the Netherlands, Hulet has calculated that the temperature below which lithium-6 — a fermion — will pair up is close to what can already be achieved by laser cooling⁶. With a little push from an applied magnetic field, it might form a

fermion condensate, and Hulet believes his team may be just a few months away from achieving this.

But several researchers, including Hulet, remain unsure about whether the theory behind these predictions is correct. With this in mind, Peter Zoller of the University of Innsbruck in Austria and his collaborators have suggested another method for generating a fermion condensate. They propose to use a grid of criss-crossing lasers. At low laser power, the atoms' wavefunctions would be spread out across the whole grid, and at high power, the atoms should be contained within the laser lattice like eggs in an egg-box. But in unpublished work, Zoller's group has shown that, in theory, adjusting the laser power to an intermediate level can cause wavefunctions to match up and atoms to form pairs.

Joint efforts

The creation of a fermion condensate would present quantum physicists with exciting opportunities. Unlike metal superconductors, dilute gases are transparent, so researchers can see into the condensate. Fermion condensates are also interesting because all the fundamental building blocks of matter — from quarks to electrons, protons and neutrons — are fermions. So it might be possible to use atomic fermion condensates to mimic the behaviour of such particles in extreme environments, such as in neutron stars, as well as in less fraught situations, such as electrons in metals.

The condensates will also provide a platform for studying poorly understood aspects of quantum theory. Using quantum mechanics to describe simple systems, such as the hydrogen atom, is relatively straightforward. But for more complex situations, such as interacting molecules, the calculations become fiendishly difficult.

Different approaches to solving these calculations exist, but it is unclear which of these are useful. "The theory is controversial," admits Stoof. "It is not immediately

clear what is the correct approach." If they have access to both BECs and fermion condensates, and can tweak atomic interactions using magnetic fields, physicists should be able to test the theory of quantum mechanics as never before. In effect, they will be able to create 'designer gases', tailoring the properties to suit their experiments. "The ability to tune the interactions is the key advantage," says Zoran Hadzibabic, a member of Ketterle's team.

Fermion condensates could also have practical applications, including quantum computing. When the atoms in Zoller's optical egg-box interact, their wavefunctions can combine in many different ways. Quantum mechanics says that the atoms actually exist in a 'superposition' of all these different possibilities until the system is observed. If the atoms could be manipulated in the right way, each different possibility could represent a calculation, allowing the system to perform a large number of calculations in parallel.

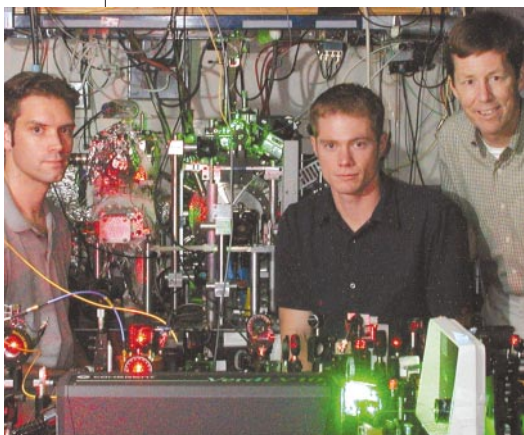
This has some potential advantages. One example is finding the factors of very large numbers, a problem at the heart of computer security and some kinds of cryptography. A hypothetical factorization calculation that would occupy a conventional computer for billions of years could be performed in months by a quantum computer.

Zoller predicts that bosons or fermions in his optical egg-box could act as bits of data in a quantum computer if they could be manipulated in the appropriate way. This January, Immanuel Bloch's group at the Max-Planck-Institute for Quantum Optics in Garching, Germany, took a very preliminary step towards this goal by applying Zoller's ideas to bosons⁷. The team placed rubidium atoms in an optical egg-box and, by varying the laser power, managed to toggle the atoms between a BEC — with full quantum behaviour — and a set of isolated atoms. Such a mechanism could be used to switch a quantum calculation on and off.

The theoretical and experimental interest in fermion condensates is surely enough to drive the groups trying to create them onto their goal. But if they need any further motivation, there is also the knowledge that fame may await; after all, those who performed the first experiments with BECs — Wieman, Cornell and Ketterle — won last year's Nobel Prize in Physics. ■

Mark Haw is a freelance science writer and a physicist at the University of Edinburgh.

1. Anderson, M. H., Enscher, J. R., Matthews, M. R., Wieman, C. E. & Cornell, E. A. *Science* **269**, 198–201 (1995).
2. Anglin, J. R. & Ketterle, W. *Nature* **416**, 211–218 (2002).
3. Ball, P. *Nature* **411**, 628–630 (2001).
4. Liu, C., Dutton, Z., Behroozi, C. H. & Hau, L. V. *Nature* **409**, 490–493 (2001).
5. Donley, E. A., Claussen, N. R., Thompson, S. T. & Wieman, C. E. *Nature* **417**, 529–533 (2002).
6. Stoof, H. T. C., Houbiers, M., Sackett, C. A. & Hulet, R. G. *Phys. Rev. Lett.* **76**, 10–13 (1996).
7. Greiner, M., Mandel, O., Esslinger, T., Hänsch, T. W. & Bloch, I. *Nature* **415**, 39–44 (2002).



Randall Hulet (far right) and his team hope their apparatus will produce a fermion condensate.

The society of proteins

Having realized that proteins usually do their jobs by combining to form transient complexes, biologists are queuing up to study these structures using a powerful electron-microscopy technique. Alison Abbott reports.

Giulio Superti-Furga likes to call himself a protein sociologist. His team has provided powerful evidence for an emerging biological concept: that proteins generally don't work alone, but instead assemble into complexes until their job is done. This can mean a long-lasting association, but often it is just a fleeting alliance. "There are two types of protein — those that prefer long-term relationships, and those that prefer one-night stands," quips Superti-Furga, who is vice-president of Cellzome, a biotech spin-off from the European Molecular Biology Laboratory (EMBL) in Heidelberg, Germany.

Behind Superti-Furga's humour lies an important shift in biological thinking. It has long been known that some proteins form stable complexes to perform a specific task. The ribosome, for instance, which manufactures new proteins from amino acids, consists of some 50 proteins plus three large RNA molecules. But the classic view of many cellular processes involves proteins interacting with one another in linear pathways, coming together as they shift around in the cell's cytoplasm.

In recent years, however, biologists have realized that many important cellular functions are actually carried out by protein complexes that act as molecular 'machines'¹.



Hunted down: Giulio Superti-Furga's team has identified more than 200 protein complexes in yeast, here represented by coloured bubbles.



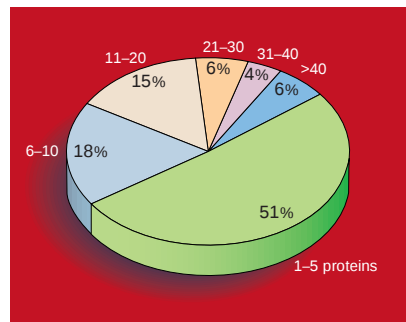
About a dozen such complexes have already been subjected to intense investigation. And structural biologists recently started to get a handle on what some of these machines look like — and how the individual proteins that form their working parts move — using new methods in electron microscopy.

It was a pair of papers^{2,3} published in *Nature* in January, one from Superti-Furga's group at Cellzome and the EMBL, the other from a Danish-Canadian collaboration headed by Mike Tyers of the University of Toronto, that really opened biologists' eyes to the ubiquity of protein complexes. The two teams tagged individual proteins from the yeast *Saccharomyces cerevisiae* with peptides that could be recognized by specific antibodies. This allowed the proteins to be used as 'bait' to pull out other proteins that were associated with them from yeast cell extracts. The researchers then purified the captured complexes and identified their constituent proteins using mass spectrometry.

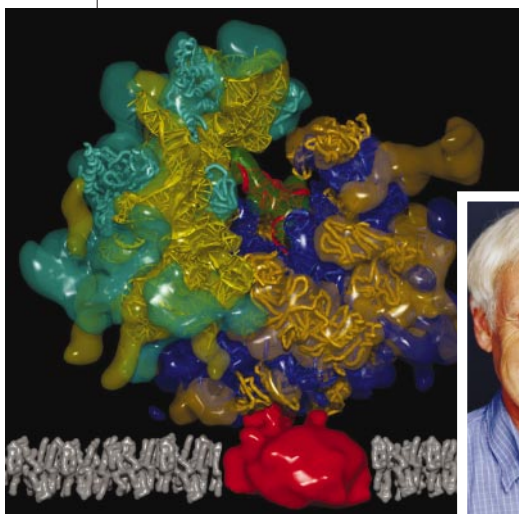
Both teams found surprising numbers of protein complexes. The analysis by Superti-

Furga and his colleagues used some 1,700 bait proteins, representing almost 30% of the yeast's total complement of expressed genes, and defined 232 distinct complexes. Superti-Furga is convinced that most are genuine functional entities, rather than experimental artefacts. "We checked that each component is found in the same subcellular compartment as the others in the same complex," he says. "We also pulled out the same complexes using different component proteins as bait."

The complexes vary considerably in size (see chart, below). Some are familiar —



Great and small: the complexes found in yeast vary in the numbers of proteins they contain.



Joachim Frank used his fast-freeze electron-microscopy technique to analyse the ribosome.



CELLZOME

CELLZOME

SOURCE: REF. 2

Arp2/3, for instance, is an assembly of seven proteins that causes the polymerization of actin, a protein that interacts with myosin to generate cell movement and make muscles contract. The structure of this complex had been published only a few months before⁴.

But most of the complexes were previously unknown. Nevertheless, by looking at the proteins within each, the researchers were able to infer possible functions for some of them. As a result, the Cellzome/EMBL team suggested new cellular functions for 344 proteins, including 231 for which there was previously no information.

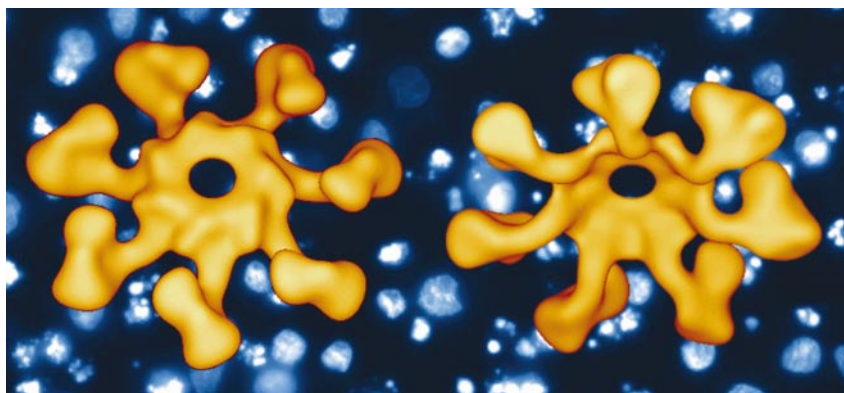
Promiscuous partners

Superti-Furga believes that this rich harvest of new complexes will include many transient structures whose component proteins are one-night-standers, “and also very promiscuous”. Indeed, almost 40% of the proteins identified in Superti-Furga’s project take part in more than one complex. “Everything seems to be interconnected,” concurs Tyers, “and the most exciting thing is the number of unanticipated connections we have found.”

To leading cell biologists, this makes sense. Tony Pawson of the Samuel Lunenfeld Research Institute at Mount Sinai Hospital in Toronto, a member of Tyers’s collaboration, points out that many cell-signalling pathways share proteins. “We also know that the same protein can do different things in different cellular states,” he adds.

One example is cytochrome *c*, a key player in respiration. This protein is usually anchored to the inner membranes of organelles called mitochondria, the main function of which is to generate adenosine triphosphate, the energy currency of the cell. But when the cell receives a signal to undergo apoptosis — or ‘cell suicide’ — cytochrome *c* leaks into the cytoplasm. Here, it associates with another protein called Apaf-1 to create the apoptosome, a complex that activates the enzyme procaspase-9, which breaks up the cell’s components for recycling⁵.

The apoptosome is one of a series of protein complexes whose three-dimensional structures have been sketched out using a technique called cryo-electron microscopy, or cryo-EM. In particular, a method called single-particle analysis, which was developed by Joachim Frank of the Wadsworth Center and the State University of New York, both in Albany, allows protein complexes to be analysed in three-dimensions from a set of electron-microscope images⁶. It does not deliver the high resolution of X-ray crystallography, the traditional method for analysing whole protein structures, but its main advantage is that it doesn’t require the proteins to be in crystalline form. This is key, because protein complexes, which are notoriously difficult to crystallize.



Wheel of death: the apoptosome’s characteristic seven-spoked shape has now been visualized.

Put simply, Frank’s technique involves applying a relatively pure solution containing the complex as a thin film onto a standard grid used to mount samples for EM, and then freezing it very rapidly at the temperature of liquid nitrogen. This freezing is so fast that the molecules have no time to move, and are captured in every possible orientation.

Electron microscopes work by shooting electron beams through a sample to generate a two-dimensional image. The complexes in the frozen film are so sensitive that they are destroyed if more than a single shot of electrons is used. But fortunately, one shot is all that is required, provided that there are sufficient complexes present, and you have enough computing power to calculate the typical three-dimensional structure of the complex from the resulting series of two-dimensional images.

All mapped out

Data on 5,000–10,000 copies of a complex are usually enough to produce a three-dimensional map of its structure with a resolution of 15–30 ångströms. Getting down to 10 Å requires up to 100,000 complexes. By comparison, X-ray crystallography can produce structures with a resolution of 2–3 Å. If these are available for any of the complex’s component proteins, however, they can be superimposed onto the lower-resolution cryo-EM map.

Cryo-EM has already helped to reveal the workings of the apoptosome. “The ability to see what it looks like in three dimensions is enormously helpful in learning about function,” says Christopher Akey, a structural biologist at Boston University who led a group that in February published the apoptosome’s structure at a resolution of 27 Å (ref. 7). Akey’s team showed that seven molecules each of Apaf-1 and cytochrome *c* organize themselves into a shape that looks like a seven-spoked wheel. This “wheel of death”, as Akey calls it, does not spin. Instead, it appears to serve as a scaffold for seven molecules of procaspase-9 to bind to its hub. The molecules would then extend from the hub into solution, allowing each to bind to



Cryo-electron microscopy provides a rapid means of unlocking protein-complex structures.

another procaspase-9 molecule to form the pairs that constitute the enzyme’s active form.

Perhaps the best demonstration of the power of cryo-EM in understanding biological processes comes from Frank’s own detailed analysis of the ribosome. A ribosome consists of two individual subunits that join together when the smaller subunit encounters and binds to a messenger RNA (mRNA) carrying the instructions to make a particular protein. Amino acids are joined together on the ribosome according to the mRNA code, each being delivered in sequence by a specific transfer RNA (tRNA) molecule. In 1995, Frank’s team first determined the shape of the empty space between the joined subunits in their cryo-EM maps⁸. “We were struck by just how similar it was to the shape of a tRNA molecule,” says Frank. “The perfection of the fit convinced us that the tRNAs must pass through this space to do their job.”

Equally revealing was the fact that the



At CIMBio (main picture), Ron Milligan hopes that cryo-electron microscopy can be automated.

shape of this space is the same in ribosomes from all sorts of organisms — bacteria, wheat, yeast, rats — despite the fact that the ribosomes themselves in these organisms are very different in composition⁹. This indicates that ribosomes must have evolved in partnership with tRNAs over the past billion years or so, says Frank. And it makes a bigger point that is beginning to dawn on biologists — that the overall shapes of protein complexes doing a particular job are more likely to be conserved across species than those of the individual protein components.

Shopping spree

Biologists worldwide are now clamouring to purchase cryo-EM equipment, and to hire experts in the technique — still a rare breed — so that they can begin to analyse a multitude of complexes.

With this in mind, Superti-Furga and his EMBL colleagues are planning to launch a project called EMBA-C (European Molecular Biology Alliance for the analysis of Protein Complexes) to coordinate efforts to generate structural information on protein complexes, and to collate and archive the data. EMBL research groups will have first pick of the protein complexes identified by the Cellzome/EMBL team. “But we hope that the alliance will eventually be open to everyone,” says Luis Serrano, who coordinates the structures and biocomputing programme at the EMBL. The effort will also extend to protein complexes identified by other groups.

Across the Atlantic, Ron Milligan, director of the newly opened Center for Integrative Molecular Biosciences (CIMBio) at the Scripps Research Institute in La Jolla, California, hopes to speed the rate of

progress by automating cryo-EM. “I was extremely excited when I read the papers in *Nature*,” he says. “Structural analysis of previously unknown complexes is exactly what we want to do with the new technologies being developed at CIMBio.”

“EM structure determination is a bit of a craft at the moment,” Milligan explains. “It involves sitting for hours at the microscope, and manually processing data.” Together with his colleagues Bridget Carragher and Clint Potter, Milligan has proved in principle that the picture taking, data processing and generation of the three-dimensional map can all be automated, using the structure of the tobacco mosaic virus as an example. “But we are a long way from generalizing this automation to all samples,” Milligan warns.

Other groups are aiming to focus on particular classes of protein complex. The Structural Cell Biology of DNA Repair Machines collaboration, led by John Tainer of the Scripps Research Institute and Priscilla Cooper of the Lawrence Berkeley National Laboratory in California, kicked off this year with a five-year grant from the National Cancer Institute. It aims to study the transient protein complexes that form when proteins that monitor DNA for damage find a problem.

Quick fix

According to the type of damage involved, particular combinations of DNA-repair enzymes join forces to achieve the necessary fix. “These transient machines are only created under particular conditions, and they are made up of parts which can recombine with other complexes to form different machines,” says Eva Nogales, a

cryo-EM expert at the University of California, Berkeley, and the Lawrence Berkeley National Laboratory, who is a member of the collaboration.

Milligan and Steve Almo, a protein crystallographer at the Albert Einstein College of Medicine in New York, this month launched discussions with about a dozen colleagues from different labs to work out a strategy for generating new protein complexes on a large scale and analysing their structures — and for seeking funding. They are also talking with Cellzome and the EMBL about creating an international ‘complexomics’ consortium. “Cellzome’s resources could serve as a foundation for getting such a programme on its feet,” says Almo. “A synergistic international effort will maximize everyone’s productivity — and fun.”

One top priority for many researchers is extending work on complexes to human proteins. Cellzome’s scientists, for instance, are already looking at pathways in different human cells, such as liver cells and those of the immune system, that are likely to be involved in human diseases. “We have shown that many complexes in yeast and humans are nearly identical — they seem to share the same sociology, as if they are conserved at the machine rather than the protein level,” says Superti-Furga. So Cellzome has been tagging all genes that have been shown to be associated with a particular disease, such as Alzheimer’s or specific cancers. In this way, Cellzome researchers hope to identify machines, rather than proteins, that can serve as a targets for drug development.

As more complexes are identified and analysed, biologists are anticipating fundamental insights into the detailed workings of the cell that could also advance our understanding of development and evolution at the molecular level. Pawson, for instance, suspects that structural information on transient complexes may explain how different growth-factor receptors manage to transmit their own particular message to the cell’s nucleus despite using the same signalling pathway.

Whatever the new discipline of analysing protein complexes ends up being called — complexomics, protein sociology, or some yet-to-be-coined term — it seems certain to bloom over the next few years. “Half of my lab wants to stop their present projects and work on complexes,” says Almo. ■

Alison Abbott is *Nature*’s senior European correspondent.

1. Alberts, B. *Cell* **92**, 291–294 (1998).
2. Gavin, A.-C. *et al.* *Nature* **415**, 141–147 (2002).
3. Ho, Y. *et al.* *Nature* **415**, 180–183 (2002).
4. Robinson, R. C. *et al.* *Science* **294**, 1679–1684 (2001).
5. Zou, H., Henzel, W. J., Liu, X., Lutschg, A. & Wang, X. *Cell* **90**, 405–413 (1997).
6. Frank, J. *BioEssays* **23**, 725–732 (2001).
7. Acehan, D. A. *et al.* *Mol. Cell* **9**, 423–432 (2002).
8. Frank, J. *et al.* *Nature* **376**, 441–444 (1995).
9. Spahn, C. M. T. *et al.* *Cell* **107**, 373–386 (2001).

Conflicts around a study of Mexican crops

Nature published a paper last year claiming that transgenic DNA had become genetically incorporated into traditional maize in Mexico. A debate ensued ...

Sir — Scientific endeavour is based on formation and testing of hypotheses. Very few hypotheses persist unmodified after they are first proposed; most are tested, modified and re-tested, and often refuted. Many journals include a forum for scientists to make technical comments on recent publications and for the original authors to respond, so that readers can evaluate the merit of reported scientific findings.

The cornerstones of a rigorous publication process are the subject editor, who is familiar with the research area of a submitted manuscript, and the independent outside reviewers whose recommendations are solicited by the subject editor. This thorough evaluation ensures, to some extent, the journal's impartiality in publication decisions.

Despite this rigorous process, however, *Nature* recently published a technical exchange^{1–3} accompanied by an editorial note stating: “*Nature* has concluded that the evidence available is not sufficient to justify the publication of the original paper”.

In our view this statement reflects poorly on *Nature*'s editorial policy and review process, and sets a dangerous precedent. Why has *Nature* refrained from releasing similar editorial retractions of earlier publications later found to be incorrect or open to alternative interpretations? What sets this particular publication apart? If the interpretation of the results proposed by the authors of the original paper⁴ was judged by *Nature* to be sufficiently erroneous to warrant this editorial statement, why did *Nature* publish the report in the first place?

By taking sides in such an unambiguous manner, *Nature* risks losing its impartial and professional status. This is particularly troubling when articles are related to economic or political interests. *Nature* asks its contributors to provide information regarding conflicts of interest, but does *Nature* hold itself to the same standards?

Andrew V. Suarez

Department of Environmental Science, Policy and Management, Division of Insect Biology, University of California, 201 Wellman Hall, Berkeley, California 94720-3112, USA

Other signatories to this letter:

Mike Benard, **Neil D. Tsutsui** Department of Evolution and Ecology, Center for Population Biology, University of California, Davis, USA

Todd A. Blackledge Department of Entomology, Cornell University, Ithaca, USA

Kirsten Copren, **Eli M. Sarnat**, **Alex L. Wild** Department of Entomology, University of California, Davis, USA

Wayne M. Getz, **Philip T. Starks**, **Kipling Will**, **Per J. Palsbøll** Department of Environmental Science, Policy and Management,

University of California, Berkeley, USA

Mark E. Hauber Department of Neurobiology and Behavior, Cornell University, USA

Craig Moritz Department of Integrative Biology, University of California, Berkeley, USA

Adam D. Richman Department of Plant Science and Plant Pathology, Montana State University, USA

Sir — The controversy surrounding Quist and Chapela's findings of transgenic introgression in Mexican maize^{1–4} is taking place within webs of political and financial influence that compromise the objectivity of their critics.

The eight authors of the two published criticisms^{1,2} of Quist and Chapela's paper⁴ have had all or part of their research funded by the Torrey Mesa Research Institute (TMRI), an offspring of the agricultural biotechnology company Novartis (now Syngenta).

The affiliation of seven of those authors with TMRI is a result of that company's \$25-million 'strategic alliance' with the University of California (UC), Berkeley's College of Natural Resources. Wilhelm Gruissem, formerly of UC, Berkeley and architect of the strategic alliance, whose laboratory is in partnership with TMRI, is the supervisor of the eighth author, Johannes Fütterer. None of the eight authors declares this funding as a competing financial interest in their published contributions.

Such a funding arrangement might be less noteworthy had Chapela and Quist not been leading critics of the strategic alliance and its implications for scientific freedom and balance. Their vocal opposition to the alliance jeopardized a large flow of financial support for the authors of the criticisms.

Compromised positions extend beyond those of these critics. Nature Publishing Group actively integrates its interests with those of companies invested in agricultural and other biotechnology, such as Novartis, AstraZeneca and other 'sponsorship clients', soliciting them to “promote their corporate image by aligning their brand with the highly respected Nature brand” (see <http://npg.nature.com>). These partnerships seem to us to challenge *Nature*'s ability to provide a neutral forum for scientific debates on agricultural biotechnology.

Nature's editorial note disavowing Quist and Chapela's work in the same issue as the technical exchanges^{1–3} was unorthodox and unnecessary. The usual scientific process of contestation should have been permitted to proceed, using

Quist and Chapela's claims and data alone to repeat, verify or refute their findings. Because of its potential effect on regulatory policy, publication of the technical exchanges and *Nature*'s editorial note immediately before the UNEP Convention on Biological Diversity meeting and discussions of the Cartagena Protocol on Biosafety (where *Nature*'s editorial note was mentioned) further undermines the journal's stance as uncompromised by commercial interests.

In such an environment, it is difficult to imagine fair consideration being given to work that challenges commercially vested interests and the assumptions of reductionist molecular biology. Quist and Chapela's paper obviously represents such a challenge. That fact — not the quality of their work — together with the politics of university–industry relations, remains central to their paper's troubled reception.

The agricultural biotechnology industry undermines its own credibility by not aggressively evaluating the health and environmental implications of its products. The public will remain sceptical until it does so.

We call on scientists, *Nature* and other scientific journals to re-examine their commitment to agricultural biotechnology as well as their own conflicts of interest, and to actively encourage a balanced, critical evaluation of the ecological and health effects of the flow of transgenes into the environment.

For further analysis, see www.cnr.berkeley.edu/~kenw/maize/

compromised.htm.

Kenneth Worthy*, **Richard C. Strohman†**, **Paul R. Billings‡**

*Department of Environmental Science, Policy and Management, University of California, Berkeley, 135 Giannini Hall, Berkeley, California 94720-3312, USA

kworthy@nature.berkeley.edu

†Department of Cell and Molecular Biology, University of California, Berkeley, 229 Stanley Hall, Berkeley, California 94720-3206, USA

‡GeneSage, Inc., 589 Howard Street, San Francisco, California 94105, USA

Statement of competing financial interests: the authors are recipients of educational grants from and/or employees of the University of California, Berkeley, which could lose financially from publication of this Correspondence as a result of its strategic alliance with TMRI/Syngenta.

Metz and Fütterer reply — The accusations of Worthy *et al.* about our Brief Communication¹ do not relate to the scientific data of Quist and Chapela⁴. Our

concern was exclusively over the quality of the scientific data and conclusions, which would have been the same whatever the motivation of the criticism.

Bad science can only undermine our understanding of nature, and the making of constructive public policy. The statement that “commercially vested interests” (that is, ties to Syngenta/Novartis) are “central to” criticisms of the data in ref. 4 is as useless in addressing the scientific issues as would be an accusation that these data were tainted by a grudge between Chapela and his former employer (the same company). Although our connections to industry are irrelevant to the scientific issues, and hence do not warrant disclosure, we feel compelled to dispel the mischaracterization of Worthy *et al.* One of us (M. M.) had TMRI funding for only one-sixth of his study at UC Berkeley, and the other’s (J. F.) alleged link to TMRI relies entirely on someone else’s former Berkeley association. Both of us currently have research funding exclusively from the public sector.

We are not unlike many scientists in that we have shared research and funding with industry at some point. In stating that we have “compromised positions”, Worthy *et al.* wrongly imply that private-sector funding strips us of integrity and legitimacy in the arena of scientific discourse. Far from promoting “scientific freedom and balance”,

this presumption tars any scientist who can be suggested to have worked with the private sector. The only threat to academic freedom that seems to have materialized from the Berkeley/TMRI collaboration is this attitude towards scientists who might have industry links.

Matthew Metz*
Johannes Fütterer†

*Department of Microbiology, University of Washington, Seattle, Washington 98195, USA

†Institute of Plant Sciences, ETH, CH-8092 Zürich, Switzerland

Kaplinsky replies — I, on behalf of the authors of our Brief Communication², state unequivocally that funding from TMRI has absolutely nothing to do with our criticisms. Worthy and co-authors are incorrect. Two of my co-authors of ref. 2 (Hake and Hay) do not receive any industry funding. Funding information for the Freeling lab (Braun, Freeling, Lisch and N. K.) is transparent and public (see <http://plantbio.berkeley.edu/~freeling/labweb/fund.html>); less than a quarter of it is from industry.

As Worthy *et al.* state, Chapela and Quist are “leading critics” of the TMRI agreement. Chapela is a board member of PANNA (<http://www.panna.org/panna/about/board.html#ihc>), an advocacy group opposing genetically modified organisms (GMOs). It is a double standard

to accuse us, but not Quist and Chapela, of a conflict of interest.

Our letter was a critique of poorly conducted and interpreted science and was not pro- or anti-GMO or industry. We simply corrected what we think is bad science. Even if we were in the pockets of industry, Quist and Chapela’s published results⁴ would still be artefactual.

Nick Kaplinsky

Department of Plant and Microbial Biology, University of California, Berkeley, California 94720, USA

1. Metz, M. & Fütterer, J. *Nature* **416**, 600–601 (2002).

2. Kaplinsky, N. *et al.* *Nature* **416**, 601 (2002).

3. Quist, D. & Chapela, I. H. *Nature* **416**, 602 (2002).

4. Quist, D. & Chapela, I. H. *Nature* **414**, 541–543 (2001).

Nature comments:

It is highly unusual for *Nature* to publish a paper whose principal conclusion is shown to be not necessarily false but unsustainable on the basis of the reported evidence. The paper was not formally retracted by its authors or by *Nature*. In the circumstances, *Nature* considered it appropriate for the record to make clear to readers its revised view of its original decision to publish.

The independence of our editorial decision-making from partisan anti- or pro-technology agendas and from commercial interests is paramount in our role as a journal. Editor, *Nature*

Impact-factor rewards affect Spanish research

Sir — In Spain, as in Finland¹, publication of research reports in journals with a high impact factor has since 1989 officially been part of the national system for evaluating researchers’ productivity. But unlike the Finnish system, the Spanish system rewards individuals rather than departments or institutions.

As stated in the Spanish parliamentary record^{2,3}, a bonus is awarded only for “those articles of scientific worth in journals of recognized prestige in the field. As a quality indicator, the relevance of the medium of dissemination in which each article was published shall be considered. In those disciplines for which international systems of quality of publications exist, reliance on these systems shall be obligatory.”

What exactly are these “systems of quality of publication”? In mathematics and physics, chemistry, cell and molecular biology, medicine, natural science, engineering and architecture, economics and social science, the law specifies that “preference shall be given to those contributions consisting of articles in journals of

recognized prestige, which shall be accepted to mean those that occupy relevant positions in the lists for science fields in the Subject Category Listings of the Journal Citation Reports of the Science Citation Index (Institute for Scientific Information [ISI], Philadelphia, PA, USA)” (ref. 4). For the first four fields, the National Commission for the Evaluation of Research Activity (CNEAI) recommends that articles should be published in journals that occupy a relevant position in the pertinent ranking, which is understood to mean the upper third of the listing ranked by impact factor. There may, however, be some flexibility under special circumstances at the discretion of the experts. The aim of this reward system is to improve the quality of Spanish science and its visibility in journals.

The consequences of this law have been first, a change in Spanish scientists’ publication habits; second, an increase in the number of Spanish source items in the ISI databases^{5,6}; and third, a levelling off of source items in Spanish bibliographic databases⁷. Spanish scientific productivity has doubled: more than twice as many papers were published between 1991 (immediately after the system of publication bonuses was passed into law)

and 1998 compared with the period 1982–1990. The acceleration in national publication output cannot be attributed to increased financial support (research budgets had begun to shrink in real terms by 1991), international collaboration or mobility of researchers, nor can it be attributed to an overall increase in scientific activity⁸.

We thank K. Shashok for translating this Correspondence into English.

Evaristo Jiménez-Contreras*, Emilio Delgado López-Cózar*, Rafael Ruiz-Pérez*, Víctor M. Fernández†

*Department of Library and Information Science, School of Library and Information Science, University of Granada, 18071 Granada, Spain

†Former director, National Commission for the Evaluation of Research Activity (CNEAI), Department of Biocatalysis, Institute of Catalysis, CSIC, 28049 Madrid, Spain

1. Adam, D. *Nature* **415**, 726–729 (2002).

2. *Boletín Oficial del Estado* 28563 (9 September 1989).

3. *Boletín Oficial del Estado* 3566 (6 February 1990); *Boletín Oficial del Estado* 37030 (3 December 1994).

4. *Boletín Oficial del Estado* 35028 (20 November 1996).

5. Garfield, E. www.garfield.library.upenn.edu/papers/maps4world.html.

6. Gómez, I., Bordons, M., Camí, J. *Science* **276**, 883–884 (1997).

7. Fernández, M. T., Bordons, M., Sancho, R., Gómez, I. *Arbor* **639**, 327–345 (1999).

8. Jiménez-Contreras, E., Moya, F., Delgado López-Cózar, E. *Research Policy* 2002 (in the press).

Our molecular history

What can DNA evidence tell us about human evolution?



BRIDGEMAN ART LIBRARY

Where Do We Come From? The Molecular Evidence for Human Descent

by Jan Klein & Naoyuki Takahata
 Springer: 2002. 462 pp. £31.50, \$49.95
 Masatoshi Nei

Jan Klein is a renowned immunologist and one of the founders of modern immunogenetics. Over the past 20 years his research has diverged to the molecular evolution of the major histocompatibility complex (MHC) and other genetic systems, and he has made significant contributions to evolutionary biology. He has written more than 500 scientific papers and authored or edited 10 books, mainly on immunogenetics.

His new book is co-authored by Naoyuki Takahata, a theoretical population geneticist who has collaborated with Klein during the past decade or so to study the molecular evolution of MHC genes. It tells the story of human descent on Earth over the past four billion years, along with some speculation about future evolution. Although it focuses mainly on molecular evidence rather than fossil evidence, the latter is covered in sufficient detail.

The authors say they wrote this book for those who want to be informed rather than entertained, and warn that it will be heavy going in places because some arguments are presented in mathematical forms. But there are enough literary references and personal references to major historical figures in the field to keep the reader's interest. The authors skilfully merge science with novels, poems, art and music — even the titles of chapters and sections are poetic.

The book starts with a chapter discussing the meaning of Paul Gauguin's painting

Where Do We Come From? What Are We Going? (shown above), from which the book takes its title. Equating "we" to the human species, they then set out to answer these questions, with emphasis on the first, which is deemed to be the most basic. Their plan is grandiose: to explain the descent of humans from primitive bacteria to modern human populations, using DNA archives. In the next four chapters the authors explain molecular techniques for extracting and sequencing DNA and statistical methods for reconstructing phylogenetic trees. The explanation starts with the basics of molecular biology and biotechnology, as well as simple population genetics and molecular phylogenetics. The explanation is lucid, but many details are left out because space is limited.

The story of opening up the DNA archives to reveal human history occupies chapters 6 to 10. Chapter 6 concentrates on the tree of life, with a detailed discussion of three domains of life. Finishing this story, the authors then move on to the evolution of metazoans, mammals, primates and, finally, to the identification of the evolutionary position of humans. Molecular data are presented with a healthy mix of palaeontological data. The authors resist anthropocentrism, and present an interesting argument for the supremacy of the cockroach. The evolution of hominids and human populations, especially the 'out of Africa' theory, is treated with special care and scrutiny.

In my view, the most interesting part of this book is chapter 11, in which the authors summarize their recent studies on the theory of speciation by the founder principle. This theory, which has been popular for the past 40 years, asserts that a new species can arise

from a small number of individuals after the population goes through a bottleneck during which genetic revolution occurs because a new set of allelic combinations are formed at different loci. This is speculation, and there has been no empirical study of this hypothesis. Klein was the first to attack this theory, using highly polymorphic MHC loci to show that humans and chimpanzees share many polymorphic alleles and therefore that humans originated from a large number of individuals — more than 1,000, he later estimated with Takahata.

The MHC study was later extended to estimate the founding population size of Darwin's finches in the Galapagos and cichlid fish in Lake Victoria and Malawi, which had been considered to be model cases of speciation by the founder principle. MHC loci again showed that the founding population size was larger than previously thought. These results led to one of the most important findings in evolutionary biology in recent years: that speciation by the founder principle may not be very common after all.

In the final chapter, entitled "What are we? Where are we going?", the authors passionately discuss several environmental issues and human affairs that should be solved urgently. There is also an appendix in which the derivation of various mathematical formulae in the text is presented.

This book contains an enormous amount of information, and anyone who masters the entire book would know a great deal about human evolution as well as molecular evolution. But some readers might find the technical details of molecular biology and mathematical formulation rather painstaking and time consuming. As a straightforward-minded theoretician,

I found the authors' approach too leisurely. I sometimes felt that I didn't need the peripherals, I just wanted to get the points more quickly, with simpler arguments. Even so, I learned many new things from this book. So my advice to the reader is: be patient and enjoy this deeply scientific and superbly artistic book. ■

Masatoshi Nei is at the Institute of Molecular Evolutionary Genetics, Pennsylvania State University, University Park, Pennsylvania 16802, USA.

Winning the fight against liver disease

Hepatitis B: The Hunt for a Killer Virus

by Baruch S. Blumberg

Princeton University Press: 2002. 272 pp.

\$27.95, £19.95

Arie J. Zuckerman

Hepatitis B is a common viral infection of the liver and a major global public-health problem. More than 2 billion people alive today — over a third of the world's population — have been infected with the hepatitis B virus, and of these about 350 million will become carriers of the virus. Some 20–25% of carriers will progress to serious liver disease, including chronic active hepatitis, cirrhosis and primary liver cancer. Primary liver cancer is the seventh most common cancer in males and the ninth most common in females, and the hepatitis B virus is second only to tobacco among the known human carcinogens. Although the precise molecular mechanisms by which the virus induces malignancy remain largely unknown, it is the cause of 80% of the world's primary liver cancer, which is one of the three most common causes of cancer deaths in males in east and Southeast Asia, the Pacific Basin and sub-Saharan Africa.

The discovery by Baruch Blumberg of the Australia antigen, a specific viral marker of the hepatitis B virus, was one of the most important advances in medical knowledge during the past 50 years and had huge implications for preventive medicine. This inspiring book is an intensely personal and interesting account of the work of Blumberg and his close associates who, after discovering the Australia antigen, continued to work on hepatitis B and devised the first-generation vaccine for this infection.

Because of this personal focus, there is relatively little mention of the contributions made by many researchers across the world and the crucial role of the World Health Organization in the unfolding story of the five different types of viral hepatitis and the implementation of universal immunization



Looking back: Baruch Blumberg's research on hepatitis B drew on earlier fieldwork in Suriname.

against hepatitis B in over 100 countries. Rates of liver cancer have fallen significantly in regions that have immunization programmes, such as Taiwan. If immunization is introduced universally, the hepatitis B virus could be eradicated within 20–50 years.

Blumberg was born in Brooklyn, New York, and attended a yeshiva, the intellectual powerhouse of traditional Jewish education. He continued his education at Far Rockaway School in New York, a public high school that counts three Nobel laureates, including Blumberg, among its alumni. He read medicine at Columbia University College of Physicians and Surgeons before leaving New York to read biochemistry at Balliol College, Oxford, UK.

The discovery of the Australia antigen was based on Blumberg's interest in genetic variation and specific susceptibility to disease, and his use of the Ouchterlony double-gel diffusion test to detect antigen–antibody interactions by the formation of precipitin lines in a gel. Blumberg did not set out to find a marker for hepatitis B — his interest was in inherited polymorphic traits. He conducted a systematic search of sera from multiply transfused patients to find antisera that could detect new antigenic polymorphisms among low-density lipoproteins and other proteins. He also used sera obtained from Australian Aborigines as part of a collaborative study.

The antibody used to find the antigen was serum obtained from New York patients with haemophilia, who had received many transfusions (and thus almost inevitably contracted hepatitis B). The rest, of course, is history, and is narrated in an eloquent and engaging style. The paper on the association between the Australia antigen and hepatitis was initially rejected by the *Annals of Internal Medicine* in 1967 because the reviewer was reluctant to risk another false claim for the identification of the elusive hepatitis virus(es). But it was published in the end, and Blumberg received the Nobel prize in Physiology or Medicine in 1976.

The book is a gem, though it is a pity about the title, which was borrowed from an overworked newspaper headline. It is essential reading for all aspiring scientists, the faceless bureaucrats who control the budgets for medical research, and devotees of the Research Assessment Exercise (it is doubtful if Blumberg's work would have attained a significant score). And it should be read by the thousands of people who work on the control and eradication of the hepatitis B virus. ■

Arie J. Zuckerman is director of the World Health Organization Collaborating Centre for Reference and Research on Viral Diseases, Royal Free and University College Medical School, Rowland Hill Street, London NW3 2PF, UK.

A philosophical liberator

Invariances: The Structure of the Objective World

by Robert Nozick

Harvard University Press: 2001. 432 pp.

\$35, £23.95, 40.30 euros

Hans-Johann Glock

Philosophers generally have one of three distinct attitudes towards natural science. Some, like Blaise Pascal and Martin Heidegger, hold science in contempt, on the grounds that philosophy, religion and the arts provide superior insights into the nature of the world and the human condition. Others, such as Bertrand Russell, Willard Van Orman Quine and most contemporary analytic philosophers, take their lead from the results and methods of science, and tend to regard philosophy as a part of, or at least continuous with, science. Philosophers of a third type also treat science with great respect but insist that philosophy is qualitatively distinct from science, because it is a second-order discipline that, rather than describing reality, reflects on the concepts used, for example, in scientific descriptions

of reality. Following Immanuel Kant, this stance can be labelled 'critical', because it allows philosophy to question scientists about anything that is conceptually unclear and about inconsistencies that might mar the expression of their factual discoveries.

The Harvard philosopher Robert Nozick is very much of the second, scientific variety. Although *Invariances* stops short of dissolving philosophy into science, it seeks to transform philosophical problems into empirical, factual ones. Its topic, the structure of the objective world, is clearly scientific, but the focus is distinctively philosophical. Nozick is less interested in exploring specific objective phenomena than in the question of what constitutes objectivity. His answer, in a nutshell, is that a phenomenon is objective to the extent to which it is invariant under a range of possible transformations.

At the same time, he deplores the typical philosophical stance according to which many statements are necessary and invariant across all possible worlds. "Too often, philosophers insist that things must be a certain way, and they make it their business to close off possibilities," writes Nozick. By contrast, he seeks to open up surprising new conceptual possibilities.

The purpose of his 'philosophical forays' is to challenge presumed necessities, and thereby our intuitions and our received conceptual framework, by invoking unheard-of possibilities raised by modern science. This challenge includes statements such as: "Water is H_2O ". Saul Kripke and Hilary Putnam have argued that the term 'water' does not refer to a substance with certain perceptible properties (such as being a colourless and odourless liquid). Rather, they say it refers in any possible world to the

substance that has the same microstructure as the stuff we call water in the actual world. Empirical scientific investigation is needed to establish what that structure is. However, in any possible world, only that which is H_2O can be water, so it is a metaphysical necessity that water is H_2O . Nozick disputes this celebrated reasoning. If there is a possible world in which H_2O is a green and smelly substance, that substance must be water. But this is not an intrinsic necessity about that world, it is 'imported' from the fact that in our actual world we use 'water' to refer to all substances with a certain microstructure.

Nozick is commendably undogmatic in his discussions of dissenting positions, and his policy of conceptual *laissez-faire* is undoubtedly attractive. But it brings with it the danger of the kind of conceptual confusions and unacknowledged terminological shifts that the critical tradition warned against. It goes without saying that scientists and scientifically minded philosophers have every right to introduce new conceptual instruments to explain new empirical findings. But if we are to understand these instruments properly, then we must clarify their connection to established (scientific or ordinary) concepts. Nozick seems to think that such explanation is unnecessary, and that new terminology can never be faulted as long as it is interesting.

But his own explorations do not always bear out this confidence. Science can have undesirable philosophical admirers — people who appeal to scientific findings and theories in an irresponsible and uncontrolled way. An example of this is the 1996 'Sokal hoax', in which the physicist Alan Sokal wrote a spoof article riddled with scientific nonsense and persuaded a leading

cultural-studies journal to publish it. Nobody could accuse Nozick of the kind of ignorance and fraudulence betokened by some postmodernist thinkers — his discussions are exceedingly well-informed, intelligent and illuminating. As a guide to how scientific findings impinge on philosophical orthodoxies in various fields, the book is highly recommended. However, it often sacrifices clarity and argumentative rigour in its quest for suggestive link-ups with science.

This is evident in Nozick's uncritical acceptance of 'evolutionary cosmology', the highly speculative and extremely metaphorical application of genetic and evolutionary terms to whole universes, and his discussion of truth and relativism. Nozick does recognize that his *laissez-faire* attitude is close to postmodern relativism, and he attempts to show that relativism is not just coherent but partly true. But his defence is flawed: postmodernists cannot avoid local charges of inconsistency by abandoning all of our received concepts and theories at once, as this will simply leave them without anything intelligible to say.

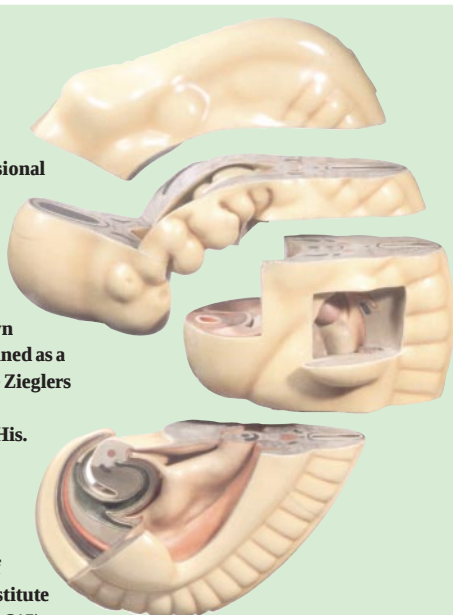
Furthermore, Nozick invokes quantum mechanics and special relativity to establish that truth is relative to both time and place (a statement is not absolutely true, but only for a certain time and in a certain place). Quantum mechanics shows that whether an event occurs at a certain time is not fixed and determined absolutely. However, there is a difference between being true and being causally determined. It is true that event *e* occurred at time *t* if, and only if, event *e* occurred at time *t*, irrespective of whether *e* was predetermined at an earlier time or measurable at a later time. Nozick simply wants us to ignore this point and to assimilate being true with what he calls 'determinately holding'. But this revision leads him to reject the most basic conceptual truth about truth: that it is true that *p* if, and only if, *p*. Anyone who rejects this equivalence is simply no longer talking about truth. (Indeed, Nozick himself relies on this equivalence when he points out that "it is true that *p*" does not mean the same as "it is believed that *p*").

Finally, Nozick demands far too much from an 'illuminating' theory of truth, because he confuses the question "What is truth?" with the question "What is true?". Answering the latter requires a scientific solution to the riddles of the universe, whereas answering the former requires 'only' an analysis of the concept of truth as used by scientists and laypeople. The role of conceptual policeman is less attractive than the role of conceptual liberator that Nozick occupies to such striking effect. But both are required if philosophy and science are to benefit from one another. ■

Hans-Johann Glock is in the Department of Philosophy, University of Reading, Reading RG6 6AA, UK.

Medical models

Wax models of embryos and other biological specimens were widely used by anatomists in the latter half of the nineteenth century. Researchers produced their own three-dimensional models to help them understand the structures they observed, and commercially prepared models were widely used in teaching. Some of the finest and most widely used wax models, such as that of a human embryo in several pieces shown here, were produced by Adolf Ziegler, who trained as a doctor of medicine, and his son Friedrich. The Zieglers acted as 'wax publishers' for scientists such as Alexander Ecker, Ernst Haeckel and Wilhelm His. The history of the Ziegler models and their impact on embryological research is told in the lavishly illustrated book *Embryos in Wax: Models from the Ziegler Studio* by Nick Hopwood (Whipple Museum of the History of Sciences, University of Cambridge, and the Institute of the History of Medicine, University of Bern, £15).



Slamming the door

John Maddox

Maxwell's 'demon', which seems on the face of it to disprove the second law of thermodynamics, has a curious place in the development of the study of energy physics. Even today, the story is surprising. One surprise is that James Clerk Maxwell himself did not spot the flaw in his own argument. After all, far from being a slouch, he was probably the outstanding scientist of the nineteenth century. A second surprise is that generations of physics students have been given the impression that Maxwell's demon is a true paradox, thereby needlessly obfuscating the whole question of the second law.

According to the historian P. M. Harman, the demon first appeared in 1867 when P. G. Tait, professor of natural philosophy at the University of Edinburgh, asked Maxwell to comment on a manuscript entitled *A Sketch of Thermodynamics* (it was published the following year). Maxwell promised to "pick some holes here and there", and duly created the demon as a means of demonstrating that the second law is statistical and therefore has a different status from 'dynamic' laws such as those of Newton.

A deeply religious man, Maxwell would not have called his device, essentially a *gedanken* (thought) experiment, a 'demon'. That term was coined by William Thomson (later Lord Kelvin). In his reply to Tait, Maxwell envisaged a vessel separated into two halves by a diaphragm, with each half containing equal numbers of gas molecules, but those on one side "having the greater energy of motion".

He continued: "Now conceive a finite being who knows the paths and velocities of all the molecules by simple inspection but who can do no work, except to open and close a door in the diaphragm,

by means of a slide without mass." The finite being's task is first to select a molecule from one half with below-average velocity and to let it through his door, and then to select a molecule from the other side with a greater-than-average velocity and to let that pass in the other direction, preserving the quantities of gas in the two halves. It will then be found that "the hot system has got hotter and the cold colder and yet no work has been done, only the intelligence of a very observant and neat fingered being has been employed".

On the face of things, this outcome may be taken as a disproof of the second law of thermodynamics in its simple form: "that heat does not flow from a colder to a hotter body without the expenditure of work," as Maxwell put it. Now, in the era of quantum physics, we know that collecting information about the velocities of all the molecules in the two compartments — by means of a Doppler measurement of scattered X-rays, for example — would entail the expenditure of work and the selective transfer of energy to individual molecules. Maxwell's demon is therefore no longer regarded as a limitation of the second law.

It is, however, also fair to ask whether Maxwell's mechanism would have been effective, in the manner he outlined, even in the classical world that he inhabited (albeit whose foundations he had already helped to weaken with the electromagnetic theory of radiation). Unfortunately, the design of the movable element of the demon's door is undefined except as a "slide", perhaps a piece of the diaphragm that separates the two volumes of gas. Presumably, it is constrained to move only in one direction by appropriate frictionless flanges, and of course the hole in the diaphragm would need to be commensurate with, but larger than, the diameter of the gas molecules.

Maxwell's demon

This famous thought experiment sets out to show a paradox in the second law of thermodynamics, but in reality Maxwell's "neat fingered being" would not get away with it.

The movable cover has one overriding function: to reflect back into the originating half of the vessel those molecules that do not satisfy the criteria of the demon's selection programme. There are two ways in which this might be achieved. Either the cover has enough inertial mass to reflect all molecules that impact upon it (in which case the demon would have to do work to move it before the calculated impact of a molecule, which contradicts the hypothesis) or it is an elastic membrane without mass.

Unfortunately, for the demon's supposed functioning, the second possibility would be subject to the constraints of the second law. First, in the normal impact of a molecule of velocity v , the molecule would first do work on the membrane of amount $0.5mv^2$, where m is the molecular mass, depositing that energy as heat. On the assumption that the elasticity of the membrane is unchanged by this dissipation of heat, the molecule would then rebound back into the container from whence it came with its original velocity.

But tensile strength is a decreasing function of temperature for all real materials, implying that molecules can be returned with their original velocities only if the heat conductivity of the membrane and of its connections to the underlying diaphragm is infinite. Nobody knows how the elastic properties of massless membranes vary with temperature, but their tensile strength cannot increase with increasing temperature, for that would be a means by which the temperature of the gas on each side of the diaphragm could spontaneously increase.

What this implies is that the functioning of Maxwell's demon must itself contradict the second law. Inevitably, it must be a device for creating negative entropy, which makes it entirely unsuitable as an element in a thought experiment that is meant to demonstrate the limitations of the second law of thermodynamics. ■

Sir John Maddox is emeritus editor of *Nature*, and is at 9 Pitt Street, London W8 4NX, UK.

FURTHER READING

Harman, P. M. *The Natural Philosophy of James Clerk Maxwell* (Cambridge Univ. Press, 1998).

VICKY ASKEW



Through a lens brightly

Edwin L. Turner

Gravitational lensing of light from some of the most distant objects known could be more widespread than had been thought. If so, it could be good — and bad — news for cosmologists.

A massive object, such as a galaxy, that happens to lie near our line of sight from Earth to a distant astronomical source can gravitationally deflect the path of light from that source. As a result, the image that reaches us may be displaced, distorted, magnified and multiplied. These cosmic optical effects, known as gravitational lensing, have been the subject of increasingly intense study by astronomers since the discovery of the first multiple images¹ of a quasi-stellar object, or quasar, in 1979. Quasars are the brightest objects in the Universe, mostly found at the centres of galaxies and thought to be associated with black holes.

Lensing can reveal valuable, and often otherwise unavailable, information about the source, the lensing object and, in cosmological contexts, even the geometrical properties of the intervening space². Because a strong lensing effect requires a rather precise chance alignment between the distant source and the foreground lensing object (Fig. 1), such systems are quite rare. For example, a few quasars in every thousand³ and only around one star in a million in nearby galaxies⁴ are strongly lensed. Gravitational-lens systems have thus been highly prized discoveries. But on page 923 of this issue, Wyithe and Loeb⁵ argue

that a substantial fraction — up to a third — of a recently discovered population of quasars are strongly lensed objects. These quasars are the most distant astronomical objects known and the light from them, emitted billions of years ago, enables us to probe the history of the very early Universe. If Wyithe and Loeb are right, their conclusions will be both good and bad news for cosmologists: some of the problems in understanding the formation and evolution of structure in the early Universe will be made easier, others more difficult.

First, the good news. The quasars, discovered by the Sloan Digital Sky Survey (SDSS)⁶, are astonishingly bright. They were discovered with exposure times of only a few minutes using a 2.5-m telescope — quite a small instrument by modern standards. A measure of how far the quasars' light has travelled to reach us is given by their redshift, z . For this distant population $z \approx 6$, which means that the Universe (and the wavelength of all photons travelling through it) has expanded by a factor of $1 + z$, so in this case a factor of seven, since the light was emitted. These high-redshift quasars are roughly as luminous as any that have been discovered at lower redshifts, and which therefore existed much later in the history of the Universe.

This is puzzling because quasars are usually believed to be powered by the rapid infall of streams of gaseous matter into super-massive black holes at the centres of their host galaxies⁷. Given their brightness, a quasar in this $z \approx 6$ population would typically be expected to contain a black hole that has a mass three billion times that of the Sun, swallowing matter at a rate approaching 100 solar masses per year! The daunting problem for theories of structure formation in the Universe is to understand how such huge black holes and the vast reservoirs of gaseous fuel were assembled so soon after the Big Bang^{8,9} — at the $z = 6$ epoch the Universe was less than a billion years old, less than a tenth of its present age.

But if strong gravitational lensing is as common in this quasar population as Wyithe and Loeb propose, the problem is eased significantly. They estimate that as many as a third of the observed $z \approx 6$ quasars may actually be at least ten times fainter than they appear, and most may be magnified by a smaller but still significant factor. The problem would not be entirely resolved though, as a quasar powered by a black hole ten times less massive and absorbing matter ten times more slowly still poses an interesting theoretical challenge.

But now the bad news. If gravitational lensing of the $z \approx 6$ quasars is common, then determining the properties of that population is more complicated and more dependent on the parameters chosen in models of the early Universe. Naïve interpretation of the quasar data, ignoring lensing effects, will lead us to overestimate not only their individual luminosities but also their total numbers and cumulative contributions to the mean radiation environment of the Universe in early times. For example, understanding the radiation environment is crucial to understanding the temperature and ionization of the diffuse gas that is found between galaxies and quasars and that is believed to have contained most of the ordinary matter in the Universe at that epoch¹⁰. Wyithe and Loeb suggest that we are seeing this early cosmic epoch modified by a complex set of gravitational lensing effects and distortions. We can attempt to correct for them, of course, but only at the cost of introducing an additional layer of systematic uncertainty and model dependence. Studies of the early Universe are

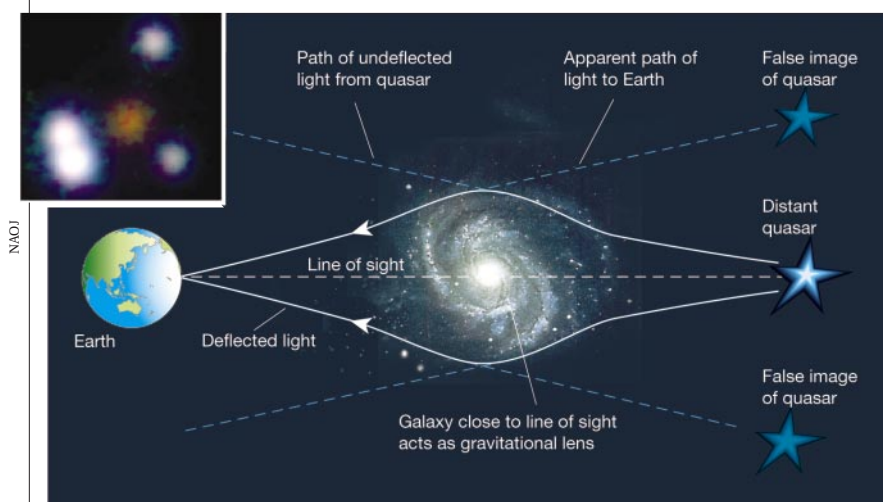


Figure 1 Gravitational lensing. Light from a distant quasar is deflected by the gravitational field of a galaxy lying close to the line of sight from Earth to the quasar. As a result, the quasar appears brighter, and multiple images of the quasar may be seen. The inset image, taken by the Japanese Subaru telescope in Hawaii, shows four images of a single distant quasar (blue) formed around the central image of an elliptical galaxy (red) that is acting as a gravitational lens. Wyithe and Loeb⁵ propose that the exceptional brightness of a population of early-Universe quasars could be due to gravitational-lensing effects that are more common than had been thought.

difficult enough without a clouded and unreliable view of the proceedings.

But why, if Wyithe and Loeb⁵ are right, should strong gravitational lensing be 30 to 100 times more common among the $z \approx 6$ quasars than among the far more numerous and extensively studied low-redshift quasars ($z \approx 2$, say)? Obviously, the $z \approx 6$ quasars are more distant, thus increasing the chance of an intervening object being near our sight line and acting as a gravitational lens. But this makes only a small contribution to the difference. The main effect is due to a selection bias — as magnification makes a strongly lensed quasar appear brighter, it also makes it more likely to be discovered. The importance of such selection biases has long been understood^{11,12} and early worries that they might compromise our ability to study the quasar population and its evolution were voiced¹² soon after the discovery of the first lensed system. But Wyithe and Loeb make a plausible case that we may only now have found the first heavily lensed population of astronomical objects.

Fortunately, a direct observational test of Wyithe and Loeb's hypothesis is possible. The strong gravitational lensing they postu-

late typically produces multiple images of a source, as well as magnifying it. These images are expected to be so close together in the sky (around 10^{-4} degrees apart) that they may not be resolvable in the SDSS images. But they should be readily resolved by other ground-based telescopes and by the Hubble Space Telescope. Such observational programmes are being initiated. We should soon have a direct measurement of the lensing that Wyithe and Loeb predict. ■

Edwin L. Turner is at the Princeton University Observatory, Princeton, New Jersey 08544-1001, USA.
e-mail: elt@astro.princeton.edu

- Walsh, D., Carswell, R. F. & Weymann, R. J. *Nature* **279**, 381–384 (1979).
- Wambsganss, J. *Living Reviews in Relativity* **1**, No. 12 (1998); www.livingreviews.org/Articles/Volume1/1998-12wamb
- King, L. J., Browne, I. W. A., Marlow, D. R., Patnaik, A. R. & Wilkinson, P. N. *Mon. Not. R. Astron. Soc.* **307**, 225–235 (1999).
- Paczynski, B. *Astrophys. J.* **304**, 1–5 (1986).
- Wyithe, J. S. B. & Loeb, A. *Nature* **417**, 923–925 (2002).
- Fan, X. *et al. Astron. J.* **122**, 2833–2849 (2001).
- Rees, M. J. *Annu. Rev. Astron. Astrophys.* **22**, 471–506 (1984).
- Turner, E. L. *Astron. J.* **101**, 5–17 (1991).
- Haiman, Z. & Loeb, A. *Astrophys. J.* **503**, 505–517 (2001).
- Madau, P., Haardt, F. & Rees, M. J. *Astrophys. J.* **514**, 648–659 (1999).
- Turner, E. L., Ostriker, J. P. & Gott, J. R. *Astrophys. J.* **284**, 1–22 (1984).
- Turner, E. L. *Astrophys. J. Lett.* **242**, L135–L139 (1980).

Cancer

Lucky draw in the gene raffle

Pamela M. Pollock and Paul S. Meltzer

The Cancer Genome Project intends to search every human gene for cancer-related mutations. Its first success is the discovery of such mutations in the *BRAF* gene.

The growth and proliferation of our cells are controlled, in part, by signals that pass from the outside of the cell to its nucleus via a cascade of molecules known as the mitogen-activated protein kinase (MAPK) pathway. Alterations in the DNA sequence of genes encoding key components of this pathway contribute to the development of many human cancers. On page 949 of this issue, Davies and colleagues¹ report that the *BRAF* gene, which codes for a protein in the MAPK cascade, is activated by mutation in some cancers — most notably in melanomas. This is the first important result to emerge from the Sanger Institute's Cancer Genome Project.

Cancer is a disease of the genome, triggered by the accumulation of genetic errors that eventually transform a normal cell into a tumour cell². Such mutations might inactivate genes that normally oppose tumour development, or activate genes that drive cell growth or interfere with cell differentiation or death. Identifying the genes that are altered in the stepwise progression to malignancy has become one of the central goals of cancer research. But finding a single disease-associ-

ated gene within the genome has been aptly compared to finding a needle in a haystack, and cancer geneticists have generally tried to

simplify the task by narrowing their search to part of the haystack before checking candidate genes for mutations.

Although the availability of the human genome sequence has not changed these unfavourable proportions, the haystack has at least been put in order, vastly accelerating the pace of disease-gene discovery. Sequencing the genome required the highly efficient automation of data generation (robotics) and computer analysis (bioinformatics)^{3,4}. Beyond producing the sequence itself, this biotechnological revolution has resulted in a significant culture shift for genome biologists, who now have these finely honed tools at their disposal.

The Cancer Genome Project has taken advantage of this technological prowess, and the human genome sequence itself, to develop a radical approach to discovering cancer-linked genes. Rather than using strategies directed at individual tumour types or chromosomal regions, Davies and colleagues have taken the bold tack of analysing the entire genome in DNA samples representing a wide spectrum of human cancers. As part of this project, they intend to examine every human gene for cancer-related (oncogenic) mutations.

Signal-transduction genes such as those of the *RAS* family, which encode components of the MAPK pathway, are frequently mutated in cancers. So Davies *et al.*¹ began their screen with other genes in this pathway, looking first at cultured cell lines derived from human tumours. They hit pay dirt with *BRAF*, which mediates growth signalling at a level just below *RAS* (Fig. 1). This gene encodes a serine/threonine kinase — an enzyme that modulates the function of other proteins by transferring phosphate groups to them. It is itself activated by such phosphorylation.

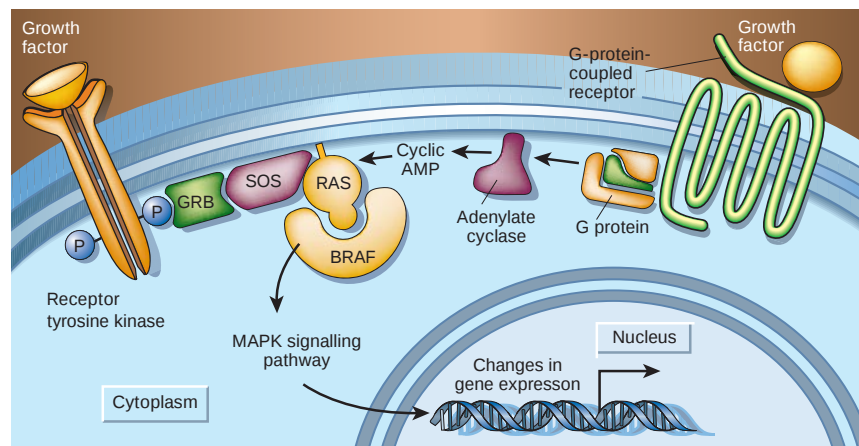


Figure 1 The *BRAF* protein and signal transduction. The *RAS*–*RAF*–*MAPK* (mitogen-activated protein kinase) signalling pathway is a pivotal molecular cascade through which extracellular signals can be transmitted into the nucleus, to control cell proliferation or differentiation by changes in gene expression. Extracellular signals (growth factors) that activate one of two types of receptor — receptor tyrosine kinases and G-protein-coupled receptors — can result in the activation of *RAS*, leading to activation of *BRAF* and the downstream cascade.

Most (80%) of the *BRAF* mutations detected by Davies *et al.* resulted in a single amino-acid change in the region of the enzyme that catalyses phosphorylation. This alteration involved the replacement of a neutral amino acid (valine at position 599) by a negatively charged one (usually glutamic acid). The mutation probably leads to constitutive activation of the MAPK pathway, by mimicking the transient phosphorylation of threonine 598 and serine 601 that occurs during normal signalling⁵. With this growth switch stuck in the 'on' position, tumour development is favoured.

Davies *et al.* found this mutation in several cancer types, mainly those known to have *RAS* mutations. For most cancers the rate of *BRAF* mutation was modest, occurring in fewer than 20% of the cell lines studied. But in melanoma cell lines, the authors found *BRAF* mutations at the staggering rate of 59%. Davies *et al.* strengthened these data by finding *BRAF* mutations in 80% of short-term melanoma cell cultures and 66% of uncultured melanomas, strongly suggesting that the high mutation frequency in cell lines was not an artefact of cell culture. This represents a clear step forward in understanding the biology of melanoma, which has so far yielded few of its genetic secrets.

Melanoma is an aggressive skin cancer that is derived from pigment cells (melanocytes). The disease can be cured surgically if caught early, but once melanoma cells have spread to other parts of the body they are resistant to most current treatments. Until now, *CDKN2A* — which encodes p16, an inhibitor of the cell-division cycle — was the most commonly mutated gene in melanoma, being inactivated by a variety of mechanisms in some 25% of sporadic melanomas⁶. Less frequent molecular alterations have been reported in *NRAS*, a member of the *RAS* family, and the tumour-suppressor gene *PTEN*.

BRAF is part of key melanocyte-specific MAPK pathways, such as that involving melanocyte-stimulating hormone, which activates the melanocortin-1 receptor on the cell surface. This may account for the high frequency of *BRAF* mutations in melanomas relative to other cancers⁷. Could *BRAF* be a breakthrough target for future treatments for melanoma? This could be so, and the intense attention that will now be directed at this gene will certainly lead to important insights. However, the identification of *RAS* mutations 20 years ago has not yet led to clinically useful *RAS*-targeted therapies.

The discovery of *BRAF* mutations¹ is a convincing validation of the Cancer Genome Project's high-throughput strategy. The mutations might not have been discovered for years without this type of approach. Now the project's investigators face the challenge of going through the rest of the genomic haystack. For their success to continue, they

must be certain that their screen is sensitive, thorough and efficient.

Challenges remain; for instance, the process of locating all genes in the human genome (annotation) still continues. Indeed, the precise number of genes is still a matter for debate, somewhat obscuring the goal of testing them all. Also, this approach depends on having DNA from cultured tumour cells; that may be difficult to achieve for some important tumours, such as prostate and pancreatic cancer, which are not easy to culture.

So where does this gargantuan effort leave the many researchers studying the genetics of cancer progression? Although it is too soon to declare focused, smaller-scale research obsolete, the Cancer Genome Project — as yet unrivalled — is likely to produce a steady flow of discoveries. Nonetheless,

the identification of a cancer-specific mutation is but the first step in a lengthy process. Further studies are needed to determine when *BRAF* mutations occur during tumour evolution, what their biochemical effects are, and the possible implications for treating cancer. ■

Pamela M. Pollock and Paul S. Meltzer are in the Cancer Genetics Branch, National Human Genome Research Institute, National Institutes of Health, 50 South Drive MSC 8000, Bethesda, Maryland 20892-8000, USA.
e-mail: pmeltzer@nhgri.nih.gov

1. Davies, H. *et al.* *Nature* **417**, 949–954 (2002); advance online publication, 9 June 2002 (doi:10.1038/nature00766).
2. Balmain, A. *Nature Rev. Cancer* **1**, 77–82 (2001).
3. Lander, E. S. *et al.* *Nature* **409**, 860–921 (2001).
4. Venter, J. C. *et al.* *Science* **291**, 1304–1351 (2001).
5. Zhang, B. H. & Guan, K. L. *EMBO J.* **19**, 5429–5439 (2000).
6. Pollock, P. M., Welch, J. & Hayward, N. K. *Cancer Res.* **61**, 1154–1161 (2001).
7. Busca, R. *et al.* *EMBO J.* **19**, 2900–2910 (2000).

Materials science

Thin-film cliffhanger

Max G. Lagally and Zhenyu Zhang

Thin films are grown ideally one atomic layer at a time, but atoms can move along and between layers. The model for film growth has now been extended to describe how atoms tumble over 'cliffs' between layers.

Many modern technologies depend on our ability to grow thin films. In highly refined semiconductor technologies (such as solid-state lasers in compact-disc players or fast electronic components in mobile phones), films must be nearly perfect, grown seemingly one atom at a time, and individual layers of a composite film may be only a few atomic layers thick. In less critical technologies, films are deposited

with essentially the same methods but with less need for perfection — protective coatings, decorative coatings, wear coatings and so on. Even something as simple as a potato-chip bag may have a thin-film coating that acts as a moisture barrier.

We think of film growth in terms of atoms or molecules 'raining' down on a surface, and then finding the proper sites to come to rest and form a layer. But in crystal growth in

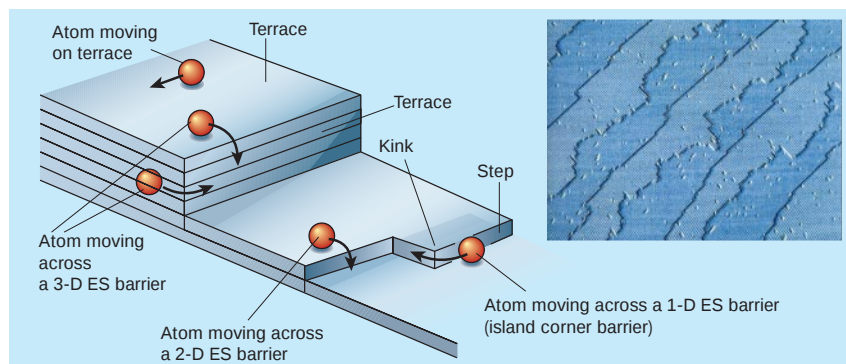


Figure 1 The terrace-step-kink (TSK) model of a thin-film surface. The surface consists of terraces separated by steps; a kink is a step on a step. Atoms travelling over steps that are one-atomic-layer high must cross an energetic barrier, the two-dimensional Ehrlich–Schwoebel (ES) barrier. At kinks, atoms experience the 'corner-crossing' barrier, a one-dimensional version of the ES barrier. Liu *et al.*¹ have identified a three-dimensional ES barrier for atoms travelling over steps that are four or more atomic layers high, or over the edges between two facets. The validity of the TSK model for thin films has been confirmed by the detailed imaging made possible by the scanning tunnelling microscope. The inset image shows the surface of a thin film of silicon (100 nm × 80 nm). Terraces separated by single-atom-high steps with many kinks can be seen, stepping down across the image from upper left to lower right. The white spots are atomic vacancies in the terraces.

general — whether it be the formation of gems over geological timescales, the appearance of salt crystals on your skin after a dip in the ocean, or the growth of alum crystals in a school science experiment — the atomic processes that occur always follow the same rules. Writing in *Applied Physics Letters*, Liu *et al.*¹ have now put in place the final piece in our understanding of the rules that govern how atoms move from one layer to another as films or crystals grow.

Consider a marble rolling off a table. It rolls over the edge, never pausing to 'think' that it might fall and break. Yet if we walk up to the edge of a cliff, we instinctively slow down, perhaps peer carefully over the edge, and draw back. There is a 'barrier' that keeps us from going forward. What would an atom do in the analogous situation?

The terrace-step-kink (TSK) model², developed by Burton, Cabrera and Frank in the early 1950s, elegantly describes the atomic-scale morphology of the surface of a crystal (Fig. 1). When an atom lands on the surface, it diffuses along the 'terrace' formed by that film layer, trying to find the best place (from the point of view of its free energy) to settle — frequently a lower terrace. But an atom coming to the edge of a terrace behaves like a human on a cliff: it feels a barrier that prevents it going over the edge.

In 1966, Gert Ehrlich performed elegant experiments³ using field-ion microscopy in which he put a single atom on an atomic terrace and observed its motion. He saw that atoms were preferentially reflected back at the edges of a terrace. He explained this behaviour in the following way. The atom on a terrace has a certain number of nearest neighbours (its coordination number) and the bonding to these neighbours provides stability for that atom. As it reaches the edge of a terrace, it suddenly has fewer neighbours, and the resulting decrease in the binding energy is manifested as a barrier for diffusion over the edge. Richard Schwoebel independently proposed⁴ a similar model at the same time.

But for more than 20 years this idea languished — until the advent of the scanning tunnelling microscope made it possible to observe the atomic-scale morphology of surfaces over mesoscopic regions, and provided vivid visual confirmation of the TSK model (Fig. 1). Because film-growth studies at the atomic level were now possible, the effects of barriers to atom transport between terraces in a growing crystal or film could be directly observed.

A reflective wall at the edge of a terrace, which became known as the Ehrlich–Schwoebel (ES) barrier, hinders the descent of atoms to lower levels. This increases the chance of nucleation and growth of a new film layer on top of the terrace when more atoms arrive from the deposition source. In 1991, Villain showed⁵ that an ES barrier (which is a two-dimensional phenomenon)

creates an effective 'uphill' gradient, leading to unstable growth and the creation of roughness in the film — an undesirable feature. Thus the two-dimensional ES barrier dictates the three-dimensional morphological evolution of growing films.

But the concept of the ES barrier also extends to other dimensions. As the figure shows, a step separates one terrace from another; a kink is the one-dimensional analogue: a 'step on a step'. We have pointed out^{6,7} that there should also be a one-dimensional analogue to the phenomenon Villain described — that an atom moving along a step edge should feel a barrier preventing it crossing this kink because of its reduced coordination number. The ultimate kinks occur at the corners of a two-dimensional crystal (an 'island'), where two step edges meet. An atom diffusing along one edge feels a barrier to crossing to the adjacent edge: this 'corner-crossing barrier' can cause two-dimensional islands (one-atomic-layer high) to develop rough, or even fractal, shapes. This corner-crossing barrier — in fact, a one-dimensional ES barrier — can also induce growth instability in the morphology of a three-dimensional film^{8,9}. Indeed, films may actually grow more smoothly if this barrier is large.

If a number of identical terraces were stacked on top of each other, the simple corner becomes a line or a ridge between two crystal facets. Following on from their earlier work¹⁰, Liu *et al.*¹ propose that there is also a three-dimensional ES barrier that influences atom transport from one facet across this edge to the adjacent facet. The authors point out that the magnitude of the three-dimensional ES barrier may be quite different from that of the two-dimensional ES barrier; the latter may, in fact, be zero but the three-dimensional barrier can be quite high.

Liu *et al.* also show that, for transitions over a step edge between adjacent terraces, the two-dimensional barrier becomes a three-dimensional barrier as the height difference between the two terraces increases (Fig. 1). For the specific example of aluminium, the transition from two-dimensional to three-dimensional ES barrier is complete if that height is four atomic layers or more^{1,10}.

So the picture is complete: the growth morphology and shape of every crystal — whether it is two-dimensional (a single-atomic-layer-high island), 'two-and-a-half-dimensional' (a nearly flat film with some roughness) or three-dimensional (a small nanocrystal) — is controlled by the magnitude of barriers felt by atoms as they approach a 'cliff' and decide whether they really do want to step off the edge. A proper accounting for these barriers in film growth will determine the ultimate quality, reliability and stability of films grown for a great variety of purposes. ■

Max G. Lagally is in the Department of Materials Science and Engineering, University of Wisconsin, Madison, Wisconsin 53706, USA.



100 YEARS AGO

Mr. Marconi's Results in Day and Night Wireless Telegraphy. Reading a brief account of these results in the *Times* of June 14, I perceive that Signor Marconi advances in explanation of the greater distance at which night signals were received, that the day signalling is affected by diselectrification of the transmitting elevated conductor. If — as I gather — Signor Marconi is referring to his observations made at positions in the Atlantic, west of England, the waves travelling westwards, may not æther drift in the earth's orbital path be concerned in producing the effects observed? The waves advancing against the orbital æther stream in the daytime, with it at night, might be supposed to give rise to conditions analogous to those which affect the transmissibility of sound against or with a high wind. It will assist if we assume a retarded æther drift near the earth's surface and free motion above. But still, the difficulty in the explanation resides in the very great magnitude of the effects observed. I write merely by way of suggestion, and in very considerable ignorance of almost every particular involved in this explanation.

J. Joly

From *Nature* 26 June 1902.

50 YEARS AGO

In the last section of a recent paper, Dirac discusses his formulæ with the following words: "An important feature of the new theory is that it involves only the ratio e/m , not e and m separately. This is what one should expect in a purely classical theory. The existence of e should be looked upon as a quantum effect, and it should appear in a theory only after quantization, and not be a property of classical electrons". If this point of view be accepted, some properties of electrons which have always been regarded as classical should be regarded as quantum effects; for example, the scattering of long electromagnetic radiation by free electrons. ... I am further convinced that it is futile to deal with the electron and its electromagnetic field separately, but that the fields of all mesons together with the electromagnetic field should be simultaneously considered. I have indicated a way of doing this elsewhere, and though I am far from thinking that this suggestion is right, I am still convinced that the solution must be sought in this direction.

Max Born

From *Nature* 28 June 1952.

e-mail: lagally@engr.wisc.edu

Zhenyu Zhang is in the Solid State Division, Oak Ridge National Laboratory, Oak Ridge, Tennessee 37831, USA.

1. Liu, S. J., Huang, H. C. & Woo, C. H. *Appl. Phys. Lett.* **80**, 3295–3297 (2002).
2. Burton, W. K., Cabrera, N. & Frank, E. C. *Phil. Trans. R. Soc. London A* **243**, 299–358 (1951).
3. Ehrlich, G. & Hudda, F. G. *J. Chem. Phys.* **44**, 1039–1049 (1966).
4. Schwoebel, R. L. & Shipsey, E. J. *J. Appl. Phys.* **37**,

3682–3686 (1966).

5. Villain, J. *J. Phys. I (Paris)* **1**, 19–42 (1991).
6. Zhang, Z. Y. & Lagally, M. G. *Science* **276**, 377–383 (1997).
7. Zhong, J. X., Zhang, T. J., Zhang, Z. Y. & Lagally, M. G. *Phys. Rev. B* **63**, 113403 (2001).
8. Pierre-Louis, O., D'Orsogna, M. R. & Einstein, T. L. *Phys. Rev. Lett.* **82**, 3661–3664 (1999).
9. Ratsch, C., Wheeler, M. C. & Gyure, M. F. *Phys. Rev. B* **62**, 12636–12639 (2000).
10. Liu, S. J., Wang, E. G., Woo, C. H. & Huang, H. C. *J. Comput.-Aided Mater. Design* **7**, 195–201 (2001).

Plant-microbe interactions

A receptor in symbiotic dialogue

Herman P. Spink

Proteins that help plants connect with symbiotic microbes have been identified. These proteins are related to receptors in animals and plants that function in the innate immune system and organ development.

Many plants can grow on soils that are poor in nutrients, and they do so by forming symbiotic associations with microbes. These associations require a molecular dialogue between the two partners. On pages 959 and 962 of this issue, Stracke *et al.*¹ and Endre *et al.*² describe how they have identified genes that encode a plant protein that is essential to the dialogue.

Important examples of microbes involved in symbiosis are bacteria called rhizobia and fungi — ‘arbuscular mycorrhizal fungi’ — that respectively supply the plant host with nitrogen- and phosphorus-containing nutrients^{3,4}. In return, the microbes receive carbohydrate nutrients from the plant. The genes identified by Stracke *et al.* and Endre *et al.* seem to belong to a large family of plant and animal genes that encode a particular class of receptor proteins. These proteins are characterized by having a repeated motif rich in the

amino acid leucine — called a leucine-rich-repeat (LRR) — in an extracellular domain^{5–7}. In animals the proteins comprise a group, known as the Toll-like receptors, that function in the innate immune system⁸. In plants they belong to a subfamily that has in common an intracellular serine/threonine kinase domain that triggers a signalling cascade inside the cell. Several members of this plant subfamily are involved in defence against pathogenic microbes, or in shoot development⁹ (Fig. 1).

It is not too surprising that receptors involved in plant-microbe symbiosis belong to the LRR receptor subfamily, given that it is one of the largest groups of receptors in the plant kingdom with, for instance, 174 members in the ‘model’ plant *Arabidopsis* alone¹⁰. But it is surprising that Stracke *et al.* and Endre *et al.* found a function for a highly similar LRR receptor in different plant species —

not only in the model legumes *Lotus japonicus*, *Medicago sativa* and *Medicago truncatula*, but also in the agriculturally important crop plant, pea (*Pisum sativum*). In all, the two groups carried out a genetic analysis of nine mutants that have defects in the early stages of symbiosis, discovering the involvement of a member of the LRR receptor family and by implication its key role in symbiosis.

For the rhizobium-plant interaction, various signal molecules produced by the microbial partner have been identified. One group — the Nod factors — has attracted much attention because Nod factors specifically trigger the host plants (belonging exclusively to the legume family) to produce a specialized microbe-accommodating organ, the root nodule³. But little is known of the plant factors that recognize rhizobial signal molecules. And even less is known of plant interactions with arbuscular mycorrhiza: for this microbe, no signal molecules have been identified.

In finding a receptor protein involved in recognizing signal molecules from both rhizobia bacteria and arbuscular mycorrhiza, Stracke *et al.*¹ and Endre *et al.*² have provided a long-awaited breakthrough. Identifying the genetic basis of microbial recognition in the plant hosts has been hampered by technical difficulties, which are especially acute for plants that form strong symbiotic relationships. As exemplified by legumes, such plants have relatively large genomes, which in some cases exceed the size of the human genome. That makes identification of point mutations — single-nucleotide changes in DNA — extremely difficult. Furthermore, in plants, gene-knockout techniques are still in their infancy^{11,12}. Stracke *et al.* and Endre *et al.* circumvented these difficulties by making optimal use of the advantages offered by classical plant genetics in legumes — the production of large numbers of mutants and genome-map-based cloning — and the increased availability of nucleotide sequence data.

The results open the way for detailed analysis of the direct binding partners, and downstream signalling pathways, that are associated with microbial recognition. Such analysis may soon provide further insight into the similarities with signal pathways involving other classes of LRR receptor, and could reveal connections with the recognition mechanism of factors produced by pathogenic microbes or signalling peptides involved in development.

The future bottleneck in studying this receptor family is our lack of knowledge about the signal molecules that are recognized by the receptors. Up to now, a triggering factor has been identified for only two plant LRR receptor proteins. These factors (microbial flagellin⁵ and the plant differentiation factor called CLAVATA3) are extracellular proteins. The structural evidence is consistent with the idea that LRR sequence

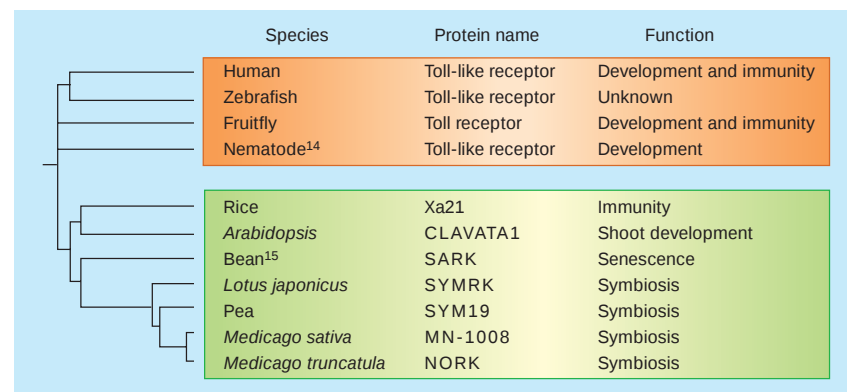


Figure 1 Comparison of members of the protein family of receptors containing extracellular leucine-rich-repeats (LRR). This tree of protein relatedness compares examples from various subfamilies in animals and plants, and indicates the breadth of species in which the receptors are found, and the variety of functions that they have. The receptors' involvement in plant-microbe interactions, as a component of the plant side of the molecular dialogue in symbiosis, is described by Stracke *et al.*¹ and Endre *et al.*² in this issue. The plants concerned are the model legumes *Lotus japonicus*, *Medicago sativa* and *Medicago truncatula*, and the pea (*Pisum sativum*). The examples not referenced above are reviewed in refs 6 and 8. For zebrafish, the assignment is based on data (BG304206) submitted to Genbank by S. Johnson. CLAVATA1 is the receptor that recognizes CLAVATA3.

motifs have a general function in protein recognition⁷, indicating that LRR receptors are only indirectly involved in recognizing carbohydrate microbial factors (such as the Nod factors). Primary recognition could be performed by secreted extracellular molecules, lectins for instance, which are recognized by the LRR receptors after binding of microbial signals¹³.

The evolutionary relatedness of the plant and animal receptors shown in Fig. 1 is surprising if one considers that the last common ancestors of higher plants and animals separated more than 1.8 billion years ago. In animals, the relatives of the plant receptors are also involved in recognizing microbes, meaning that similar carbohydrate-recognition mechanisms could underlie the animal innate immune system. Carbohydrate recognition in the human immune system is still poorly understood, so unravelling microbe-recognition mechanisms in plants should also provide lessons for medical science. ■

Herman P. Spaijk is in the Clusius Laboratory, Leiden University, Wassenaarseweg 64, 2333 AL, Leiden, The Netherlands.

e-mail: spaijk@rubiim.leidenuniv.nl

1. Stracke, S. et al. *Nature* **417**, 959–962 (2002).
2. Endre, G. et al. *Nature* **417**, 962–966 (2002).
3. Spaijk, H. P., Kondorosi, A. & Hooykaas, P. J. J. *The Rhizobiaceae: Molecular Biology of Model Plant-Associated Bacteria* (Kluwer, Dordrecht, 1998).
4. Harrison, M. J. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **50**, 361–389 (1999).
5. Asai, T. et al. *Nature* **415**, 977–983 (2002).
6. Dangl, J. L. & Jones, J. D. *Nature* **411**, 826–833 (2001).
7. Kobe, B. & Kajava, A. V. *Curr. Opin. Struct. Biol.* **11**, 725–732 (2001).
8. Schuster, J. M. & Nelson, P. S. *J. Leukocyte Biol.* **67**, 767–773 (2000).
9. Clark, S. E. *Nature Rev. Mol. Cell Biol.* **2**, 276–284 (2001).
10. The Arabidopsis Genome Initiative *Nature* **408**, 796–815 (2000).
11. Schaefer, L., Roussis, A., Stiller, J. & Stougaard, J. *Nature* **402**, 191–195 (1999).
12. Spaijk, H. P. *Nature* **402**, 135–136 (1999).
13. van der Holst, P. P. G., Schlaman, H. R. M. & Spaijk, H. P. *Curr. Opin. Struct. Biol.* **11**, 608–616 (2001).
14. Pujol, N. et al. *Curr. Biol.* **11**, 809–821 (2001).
15. Hajjouj, T., Michelis, R. & Gepstein, S. *Plant Physiol.* **124**, 1305–1314 (2000).

Earth science

How old roots lose their bounce

David James

Active mountain belts have crustal 'roots' that gravitationally balance the high topography. So why do old mountains that have been worn flat by erosion still have these roots?

Some mountain ranges form along the boundaries of colliding tectonic plates, and the classic account of their rise and fall runs as follows. In these 'collisional belts', the Earth's crust is squeezed, thickened and uplifted by the forces that drive the plates together, and is also thickened by the addition of magmas from the underlying mantle. Where powerful mountain-building forces are at work today, as in the central Andes and Tibet, the mass of the mountains is gravitationally supported by a thick 'root' of buoyant low-density rock beneath them. When active uplift and crustal thickening cease, erosion begins the inexorable process of reducing once lofty mountains to low-lying plains. As the mountains are eroded, gravitational balance is maintained by continued uplift of the buoyant crustal root to compensate for mass loss at the surface.

As Fischer reports on page 933 of this issue¹, however, this classic balancing act seems not to apply to many old and deeply eroded collisional belts. In these places, disproportionately thick crustal roots have often survived for hundreds of millions — even billions — of years, and Fischer provides an explanation for why that should be so (Fig. 1).

She has examined collisional mountain belts of all ages worldwide. She finds that the ratio of elevation to crustal-root thickness decreases systematically from about 0.1–0.2

for the youngest mountain belts (root thickness 5–10 times greater than surface relief) to essentially zero for mountain belts several hundred million years old (still appreciable root thickness, but minimal surface relief; see Fig. 1a on page 933). This is perplexing because, given the mass balance between crustal roots and surface load in young mountain ranges, it would seem that removal of the mass at the surface should involve proportional loss of the buoyant root. Old mountains eroded to flat-lying plains should have no remaining roots.

It was triangulation anomalies, observed in the Himalayas during the Trigonometrical Survey of India in the mid-nineteenth century, that led to the realization that mountains are not simply loads piled onto the surface but are compensated by a comparable 'mass deficit' at depth. This gravitational balancing act is termed isostatic compensation. In its most common form, isostasy is simply Archimedes' principle applied to the Earth: low-density crustal blocks 'float' in a higher-density 'fluid' mantle, much as an iceberg floats in water with its tip gravitationally balanced by a much greater volume of displaced water beneath the surface. Mountain ranges in isostatic equilibrium are held aloft by the buoyant forces of crustal roots, and the higher the mountains, the thicker the roots. Strictly speaking, the iceberg analogy is valid only if the crust and

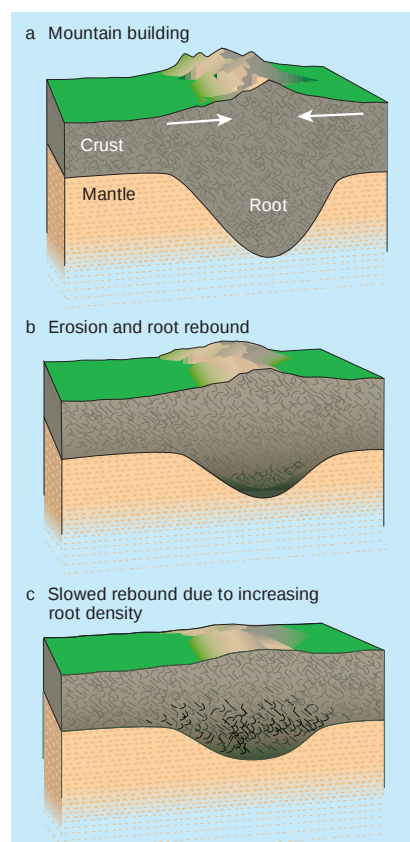


Figure 1 Mountain building at 'collisional' boundaries, and the fate of crustal roots. a, The collision of tectonic plates produces a thickened crust — a mountain range — which is gravitationally supported by an underlying crustal root of buoyant, low-density rock. b, When active mountain-building ends, the erosional loss of surface mass is gravitationally balanced by rebound of the buoyant root. The principle of isostasy suggests that continued uplift and exhumation — deep erosion — of the crust should continue until the buoyant crustal root is completely consumed. c, Fischer¹ finds, however, that older collisional belts tend to have disproportionately thick roots even when surface topography has been greatly reduced. She explains this as a result of reduced root buoyancy, caused by mineralogical reactions that increase root density relative to that of the surrounding mantle.

mantle are assumed to be sufficiently weak — deformable — that the crust really can 'float' in the dense 'fluid' mantle beneath. This assumption seems to be plausible for active mountain belts, but it might not be valid for older mountain ranges, for which progressive cooling may lead to an increase in rigidity.

Fischer assessed the role of isostatic balance by considering two hypotheses for how crustal roots could survive over geological time. The first was that the lithosphere — the crust and upper mantle — becomes increasingly rigid over time, impeding the buoyant rebound of low-density roots; the second was that the density contrast between crustal

root and mantle decreases with time. For all but the youngest mountain belts, these two hypotheses will produce very different predicted gravity anomalies, a consequence of the differing density contrast between root and mantle implicit in the two hypotheses. Fischer modelled crustal densities to obtain a best fit to the observed gravity anomalies and shows that the data do not fit the first hypothesis: old crustal roots cannot be composed of the same low-density buoyant material that makes up the roots of active mountain belts. Instead, she shows that the density contrast between the crustal root and the mantle decreases markedly with increasing age, although in almost every case the crustal root still provides isostatic compensation for the surface features.

This analysis indicates that, over time, the crustal roots must evolve in response to cooling temperatures and to mineralogical reactions that drive the lower crust to higher density. This process does not take long: root buoyancy seems to be relatively reduced for all mountain belts in which activity ended 20 million years ago or earlier. Fischer suggests that the decrease in density contrast between deep crust and mantle with age means that crustal roots remain roughly in isostatic equilibrium even as the surface is reduced to low-lying topography. Implicit in this conclusion is that the lithosphere remains weak enough for the buoyant uplift of roots to continue over long periods of geological time.

Fischer's results have broader implications, particularly as they bear upon the

widely held notion of 'gravitational collapse' of mountain belts. Gravitational collapse, also termed post-orogenic collapse, posits that the crust that is thickened during mountain building is rapidly thinned through post-orogenic 'extension'^{2,3} — stretching — which produces accelerated crustal thinning and gravitational collapse of the mountain belt itself. The process is widely assumed to be a normal stage in the mountain evolutionary cycle, one that should result in the virtual disappearance of crustal roots³. According to this view, active mountain building is followed relatively quickly by collapse of the mountain belt, driven largely by thermal erosion and thinning of the lithosphere (caused by heating from the deeper mantle), and by internal heating of the thickened crust itself.

Fischer's findings imply that post-orogenic crustal evolution is not always quite so simple. There are certainly examples of rapid gravitational collapse after mountain building. But the persistence of crustal roots in many old mountain belts, and their apparent continued uplift over hundreds of millions of years, mean that large-scale extension, thinning and destruction of crustal roots need not be the inevitable conclusion. ■

David James is in the Department of Terrestrial Magnetism, Carnegie Institution of Washington, 5241 Broad Branch Road, NW, Washington DC 20015, USA.

e-mail: james@dtm.ciw.edu

1. Fischer, K. M. *Nature* **417**, 933–936 (2002).

2. Dewey, J. F. *Tectonics* **7**, 1123–1139 (1988).

3. Vanderhaeghe, O. & Teyssier, C. *Tectonophysics* **335**, 211–228 (2001).

Genomic imprinting

Piece of cake

Benjamin Tycko and Argiris Efstratiadis

Knocking out a minor form of the *Igf2* messenger RNA from the placenta in mice has surprisingly strong effects on nutrient transport to the fetus. This has implications for the theory of maternal–paternal genetic conflict.

About 2,500 years ago, Anaxagoras (500–428 BC) was the first, or so the record indicates, to perceive that embryos receive nutriment through the navel. (Some of his contemporaries had bizarre notions, including the idea that the embryo is nourished through its whole body, like a sponge.) Soon afterwards, Diogenes of Apollonia (460–400 BC) described some of the anatomical details of the placenta — although this Latin word, which literally means 'cake' and comes from the Greek 'plakous' (flat mass), had not been coined. In the second century AD, the Roman statesman Cato the Elder left a recipe for a popular pie called placenta ('pizza' evolved much later). It is unclear whether the simile went from bakers to anatomists or the other way around, but the fact is that the placenta start-

ed receiving serious attention after the beginning of modern embryology in the late nineteenth century. Its surprising secrets continue to be revealed today, as exemplified by a paper from Constância and co-workers¹ on page 945 of this issue.

Although the details of placental structure (Fig. 1) differ between species, there is enough similarity to make data from mouse models relevant to humans². In mice, the placental layer closest to the fetus is called the labyrinth — a network of mingled fetal and maternal blood vessels, separated by a layer of so-called labyrinthine trophoblast cells. This architecture leaves little doubt that the labyrinth is crucial to the exchange of nutrients, gases and waste between maternal and fetal blood. On the maternal side of the labyrinth is a poorly understood region

called the junctional zone, composed of large spongiotrophoblast cells and another type of trophoblast cell, the glycogen cell. Next to the junctional zone are maternal decidual cells, which line the uterus. A discontinuous layer of trophoblast 'giant' cells contacts the decidual cells.

One determinant of the size of the placenta is the *Igf2* gene, which encodes the insulin-like growth factor II (IGF-II) protein. Its significance for placental growth was recognized some time ago, from gene-knockout studies in mice: the lack of IGF-II resulted in small placentas³. *Igf2* is expressed in the entire mouse placenta — in both spongiotrophoblast and labyrinth — and in the corresponding regions in humans. But in mice (not humans) one of the control regions of *Igf2*, dubbed the P0 promoter, controls labyrinth-specific expression to produce a messenger RNA copy with two small non-protein-coding regions ('mini-exons'), as well as the usual coding regions⁴. Previously⁵, Constância *et al.* described mutant mice that lacked 5,000 bases spanning the second mini-exon, eliminating the P0-derived mRNA; mutant fetuses showed retarded growth.

Constância *et al.*¹ have now studied the same mutation in more detail. They find that, although the P0 version normally makes up only 10% or so of the total placental *Igf2* mRNAs, its absence from mutants compromises the size of the entire placenta as well as the final size of the fetus. The total amount of *Igf2* mRNA is somewhat reduced in the labyrinth but not in the spongiotrophoblast. And yet all placental regions — including, amazingly, the maternal component — are proportionately reduced in size. But there is more: the placental size at the end of gestation is the same as in mice that lack *Igf2* entirely. This striking observation is difficult to interpret, unless one invokes the idea of differential translation of the various *Igf2* mRNAs into

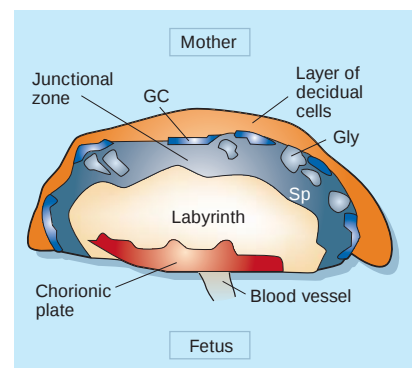


Figure 1 Cross-section of the mouse placenta. Sp, spongiotrophoblast; Gly, glycogen cells; GC, trophoblast giant cells. Fetal blood enters via the large central blood vessel. In humans, the labyrinth is known as the fetal placenta and the junctional zone is called the basal plate. Drawing modified from ref. 16.

protein⁶. This issue has yet to be addressed.

As with most genes, mammals have two copies (alleles) of *Igf2*, one from the mother and one from the father. But *Igf2* belongs to the small club of 'imprinted' genes — those that are conditioned during egg or sperm development⁷ so that only one allele is expressed in offspring. *Igf2* is imprinted in such a way that the paternal copy is expressed and the maternal allele is silenced. Given this fact, and evidence that parental-specific imprinting evolved in vertebrates along with the placenta⁸, the observations of Constância *et al.*¹ may be hinting at something profound. But what?

One hypothesis that explains the co-evolution of imprinting and the placenta incorporates the idea that there is conflict between maternal and paternal alleles. This theory predicts that, in species in which the mother, but not the father, makes a major investment in offspring, and in which multiple males sire progeny by the same female, maternal alleles will have a selective advantage if they act to limit the resources allocated to any one conceptus. In contrast, paternal alleles will gain an advantage by commandeering as many maternal resources as possible⁹. In simple terms, then, paternally expressed imprinted genes, such as *Igf2*, should promote fetal growth, and maternally expressed genes should inhibit fetal or placental growth. Data from a growing number of imprinted genes¹⁰, including some that are unrelated to *Igf2* signalling¹¹, meet these predictions. But the theory has serious difficulties with falsifiability⁷, aggravated by confusing anthropomorphisms, unbecoming militaristic metaphors ("battle of the sexes" and the like), and the vexing fact that some imprinted genes do not control growth¹².

The work by Constância *et al.*¹ not only addresses maternal–paternal conflict but also has some truly exciting implications for fetal physiology. One of the most novel aspects of the authors' work is that they took the bull by the horns and measured the passive permeability and active transport of nutrients across the mutant placentas. Using radiolabelled tracers, they showed a decrease in passive, but an increase in active, transport early in development, when normalized to placental weight. But this compensatory attempt by the stunted placentas to increase the provision of nutrients to the embryos failed. Net delivery of an amino-acid analogue, not adjusted for placental weight, was below normal.

These results imply that the paternally expressed *Igf2* acts in the placenta to channel maternal resources to the fetus; when it is mutated, other mechanisms try, but fail, to compensate. So this gene seems to score a home run for the conflict hypothesis (at least for the 10% of *Igf2* mRNA that is affected; how the other 90% fits into the picture is

unclear). Still, given that *Igf2* was one of the two genes that motivated the hypothesis in the first place⁹, there is more than a little circularity in this argument. Evidence from genetic manipulations of other imprinted genes will be grist for the mill.

Do these results relate to human health? A low birth weight correlates with health problems both soon after birth and, more perniciously, in later life¹³. The mechanism behind the late complications is still wanting, but Smith *et al.*¹⁴, as they report on page 916, have taken a step towards a simpler objective: to predict low birth weight by a practical screen. They measured levels of the pregnancy-associated plasma protein A (PAPP-A) in maternal blood in a large human population, early in pregnancy. Mothers with very low PAPP-A levels were more likely to have small babies. As PAPP-A is expressed in early trophoblast tissue, these data imply that the health and abundance of the early trophoblast correlates with birth weight. Might these data directly implicate the *Igf2* pathway? That remains to be seen. PAPP-A cleaves IGF-II-binding proteins in an IGF-II-dependent manner, but the function of these proteins *in vivo* remains uncertain. In fact, single — and some double — knockouts of the genes encoding these proteins in mice had no startling effects on growth (ref. 3 and J. Pintar, personal communication).

Normally, mice consume the placenta after giving birth (yes, they have their cake and eat it too). Like deletion of *Igf2*, and consistent with conflict, knocking out another paternally expressed imprinted gene, *Peg1*, causes fetal and placental growth retardation¹⁵. But a finding apparently unrelated to conflict is that *Peg1*-lacking mothers also show a behavioural quirk — they no longer eat the placenta. Food for thought. ■

Benjamin Tycko is at the Institute for Cancer Genetics and the Department of Pathology, and Argiris Efstratiadis is at the Institute for Cancer Genetics and the Department of Genetics and Development, Columbia University College of Physicians and Surgeons, 1150 St Nicholas Avenue, New York, New York 10032, USA.
e-mails: bt12@columbia.edu
arg@cancercenter.columbia.edu

1. Constância, M. *et al.* *Nature* **417**, 945–948 (2002).
2. Georgiades, P., Ferguson-Smith, A. C. & Burton, G. J. *Placenta* **23**, 3–19 (2002).
3. Efstratiadis, A. *Int. J. Dev. Biol.* **42**, 955–976 (1998).
4. Moore, T. *et al.* *Proc. Natl Acad. Sci. USA* **94**, 12509–12514 (1997).
5. Constância, M. *et al.* *Nature Genet.* **26**, 203–206 (2000).
6. Nielson, F. C. *et al.* *J. Biol. Chem.* **265**, 13431–13434 (1990).
7. Solter, D. *Int. J. Dev. Biol.* **42**, 951–954 (1998).
8. John, R. M. & Surani, M. A. *Cell* **101**, 585–588 (2000).
9. Haig, D. & Graham, C. *Cell* **64**, 1045–1046 (1991).
10. www.otago.ac.nz/IGC
11. Frank, D. *et al.* *Proc. Natl Acad. Sci. USA* **99**, 7490–7495 (2002).
12. Hurst, L. D. & McVean, G. T. *Curr. Opin. Genet. Dev.* **8**, 701–708 (1998).
13. Gluckman, P. D. & Harding, J. E. *Horm. Res.* **48**, 11–16 (1997).
14. Smith, G. C. S. *et al.* *Nature* **417**, 916 (2002).
15. Lefebvre, L. *et al.* *Nature Genet.* **20**, 163–169 (1998).
16. Sapin, V. *et al.* *Dev. Biol.* **191**, 29–39 (1997).

Daedalus

Darkened light

'Dayglo' dyes look brighter than they should. They absorb invisible ultraviolet, and re-emit that energy in a lower visible band, perhaps as green, orange or red. Normal dyestuffs, with a broad visible band, must also be fluorescent. Illuminated with sharp, narrow-band light, a dye will be raised to its excited electronic state, and will re-radiate that energy over its whole broad visible band. This range of re-emitted frequencies is much wider than those inserted by the narrow-band source. Luckily, the Sun and most lamps are such broad-band sources that we notice no effect.

Yet 'same-band fluorescence' should be far commoner and more efficient than the two-band variety of Dayglo dyes. So Daedalus is now lighting targets with modern narrow-band sources: lasers, metal-vapour lamps, LEDs. For each source, he has a matching pair of narrow-band rejector spectacles, whose 'notch' cuts out just the frequency of the source. Although mildly coloured, they darken the lamp itself for the wearer. But the pigment — DREADCO's Lampglo dye — fluoresces either side of that region in almost full wide-band glory. Unblinded by lamps, the wearer will see the dye very brightly.

In an ordinary room lit by broad-band tungsten bulbs, Lampglo dye will seem unexceptional. Its novel power will be felt only in rooms lit by narrow-band emitters, and to wearers of DREADCO spectacles with a coincident absorptive 'notch'. The lamps will hardly be visible, but Lampglo-dyed domestic objects will shine brightly indeed. Computer and TV screens will be quite black, but for their own light.

Daedalus plans to transform night driving as well. Sodium street lamps are narrow-band metal-vapour emitters already, so properly notched spectacles will hardly see them. DREADCO sodium headlamps on approaching cars will also fail to dazzle. But Lampglo-coloured cars will show up well, as will the Lampglo road itself.

David Jones

A rethink of the content of News & Views means that this is David Jones's last Daedalus column for *Nature*. The column has appeared for the past 14 years, apart from a five-month break at one point, and was published for 24 years before that in *New Scientist*. We owe David thanks for his consummate professionalism as a columnist, and for entertaining, stimulating, provoking and — sometimes — perplexing *Nature* readers over so many years. He can be contacted at david@davidjones.fsworld.co.uk.
Editor, *Nature*

Caffeine as a repellent for slugs and snails

At high concentrations this stimulant becomes a lethal neurotoxin to garden pests.

Most commercial products for snail and slug control contain either metaldehyde or methiocarb as the active ingredient¹, the residues of which are not permitted in food crops in the United States². We have discovered that solutions of caffeine are effective in killing or repelling slugs and snails when applied to foliage or the growing medium of plants. Because caffeine is a natural product and is classified by the US Food and Drug Administration as a GRAS ('generally recognized as safe') compound³, it has potential as an environmentally acceptable alternative toxicant for the control of slugs and snails on food crops.

While field-testing caffeine as a toxicant against an introduced frog pest that infests potted plants in Hawaii, we discovered that large slugs were killed by spray applications containing 1–2% caffeine. To test whether caffeine solutions could be used to remove or kill large slugs that attack potted plants, we allowed *Veronicella cubensis* (Pfeiffer) to bury themselves in the soil in the pots, and then thoroughly wetted the soil with a 2% caffeine solution. After 3.5 h, only 25% of the slugs remained in the soil; after 48 h, all slugs had left the soil and 92% were dead.

We tested the effect of caffeine on slugs in choice and no-choice feeding tests. Leaves of 'Napa' cabbage were dipped in caffeine solution, drip-dried and then placed in ventilated, 2-litre plastic containers (Fig. 1). In no-choice tests over a 4-day period, the weight of cabbage consumed by the slugs exposed to caffeine relative to controls was reduced by 9, 19, 29 and 39% for caffeine treatments of 0.01, 0.1, 0.5 and 2%, respectively.

Similarly, in choice tests over a 4-day period, consumption of treated cabbage leaves was reduced by –11, 62, 77 and 64% for the same concentrations. However, total cabbage consumption (that is, feeding on treated and untreated leaves) was reduced by 24, 24, 14 and 28%, respectively. We conclude that slugs can discriminate between treated and untreated leaves, and that caffeine significantly reduces feeding at a concentration of only 0.01%.

To test the effect of caffeine on snails, we treated mature orchid snails (*Zonitoides arboreus* (Say); about 3 mm in diameter) topically with caffeine solutions in Petri dishes and examined them under a dissecting microscope to determine their heart rates. One hour after treatment, the heart rates of snails exposed to 0.01% caffeine were raised, whereas those of snails treated with 0.1, 0.5 or 2% caffeine were reduced (Fig. 2a). After 24 h, heart contractions



Figure 1 Repulsion of slugs (*Veronicella cubensis* (Pfeiffer)) by caffeine-sprayed leaves. Results pictured (clockwise, from top left) are for no-choice treatments using 1.0, 0.1 and 0% caffeine, and for a choice test with 0 (left leaf) and 0.5% caffeine. Caffeine treatment causes slugs to avoid the leafy parts of the cabbage.

were weak and irregular in snails treated with caffeine concentrations of 0.1% or more, and all snails treated with 0.5 or 2% caffeine were dead after 96 h (Fig. 2b).

We then carried out two greenhouse tests in which caffeine solutions were applied to the growing medium of potted orchids infested with *Z. arboreus*. In the first test, 1 and 2% solutions of caffeine were associated with 60 and 95% mortality, respectively (control mortality averaged 10%). In the second, larger test, a 2% solution of caffeine proved to be more effective in reducing the presence of snails than a solution of 0.195% metaldehyde, the commercial standard for orchid-snail treatment. Over 30 days, 5 snails were extracted from the medium treated with caffeine, whereas 43 and 35 snails were collected from media treated with water (control) and metaldehyde, respectively.

Although caffeine has been used to study ion transport and the physiological function of nerves and muscles in molluscs and insects^{4–7}, its toxicological effects on slugs and snails have escaped attention. Caffeine has been proposed as a potential pesticide and pesticide synergist, but only with reference to insect control^{8,9}.

We do not know how caffeine kills slugs and snails. Physiological studies of mollusc neurons indicate that it releases calcium from internal stores, thereby increasing the duration of action-potential plateaus^{6,10}. We found that large slugs placed on loose soil and sprayed with a 1 or 2% solution of caffeine responded with uncoordinated writhing; the only survivors were the few that were able to burrow into the soil soon after treatment.

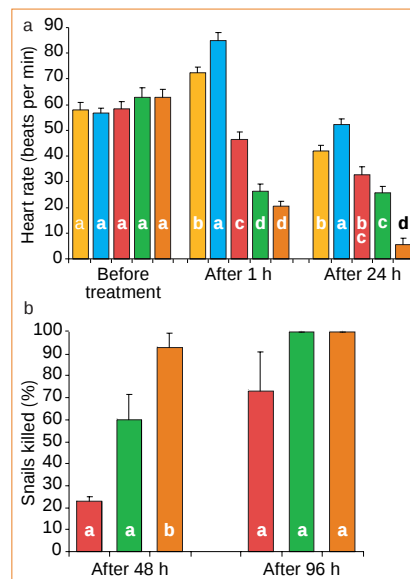


Figure 2 Effect of caffeine on the orchid snail *Zonitoides arboreus* (Say). **a**, Heart rates before and after treatment with different concentrations of caffeine. Bars (left to right in each set): 0, 0.01, 0.1, 0.5, 2.0% solutions of caffeine. **b**, Mortality induced by higher doses of caffeine. Bars (left to right in each set): 0.1, 0.5, 2% solutions of caffeine. Error bars represent 1 s.e.m. Each treatment of five snails on filter paper in a Petri dish (three dishes) involved wetting the animals and paper with 1 ml caffeine solution or distilled water (control). Heart rates of snails in the same dish were averaged before ANOVA. The proportion of snails that died in each dish were arc-sine transformed¹² before analysis, excluding control dishes and those containing 0.01% caffeine, in which no snails died. Within each time period and test, results labelled with the same letter(s) are not significantly different ($\alpha = 0.05$, Tukey's HSD; ref.13). Further details are available from the authors.

We anticipate that, in an agricultural environment, slugs and snails will be more susceptible to contact poisoning from caffeine than other animals such as arthropods. This is because caffeine is highly soluble in water¹¹, an important component of the mucus produced by the locomotive 'foot' of molluscs.

In preliminary trials, we found that 2% caffeine does not damage the foliage of *Dracaena*, *Anthurium*, palms or orchids, but that it causes leaf yellowing on ferns, bromeliads and lettuce. This damage could be ameliorated by mixing caffeine with an appropriate agricultural polymer, which could also increase the water resistance of spray residues.

Robert G. Hollingsworth*,

John W. Armstrong*, Earl Campbell††

*US Pacific Basin Agricultural Research Center, USDA-Agricultural Research Service, PO Box 4459, Hilo, Hawaii 96720, USA
e-mail: rholling@pbars.ars.usda.gov

†USDA-APHIS-WS-National Wildlife Research Center, Box 10880, Hilo, Hawaii 96721, USA
‡Present address: US Fish and Wildlife Service, Box 50088, Honolulu, Hawaii 96850, USA

1. Martin, A. *Outlook Agric.* **20**, 167–174 (1991).
2. US Code of Federal Regulations, Title 40 — *Protection of Environment* (40CFR180.523) 479 (US Government, 2001).
3. US Code of Federal Regulations, Title 21 — *Food and Drugs* (21CFR182) 455 (US Government, 2001).
4. Fahim, M. A. & Usherwood, P. N. R. *J. Neurobiol.* **14**, 391–397 (1983).
5. Moffett, D. F., Smith, C. J. & Green, J. M. *Comp. Biochem. Physiol.* **75**, 305–310 (1983).
6. Ahmed, I. A., Hopkins, P. M. & Winlow, W. *Gen. Pharmacol.* **28**, 245–250 (1997).

7. Cook, B. J. & Holman, G. M. *Comp. Biochem. Physiol.* **67**, 115–120 (1980).
8. Nathanson, J. A. *Science* **226**, 184–187 (1984).
9. Wilkins, R. M., Saleem, M. A. & Candasamy, R. *Pestic. Sci.* **43**, 321–331 (1995).
10. Ahmed, I. A., Hopkins, P. M. & Winlow, W. *Comp. Biochem. Physiol.* **105**, 363–372 (1993).
11. Windholz, M. (ed.) *The Merck Index* 9th edn (Merck, Rahway, New Jersey, 1976).
12. Steel, R. G. D. & Torrie, J. H. *Principles and Procedures of Statistics* (McGraw-Hill, New York, 1980).
13. *SAS/STAT User's Guide* 6.03 Chs 20, 24 (SAS Institute, Cary, North Carolina, 1988).

Competing financial interests: declared none.

Development

Early-pregnancy origins of low birth weight

Low birth weight is a significant cause of morbidity and mortality among newborns, and may result from impaired placental function during the first trimester of pregnancy¹. Here we show that the risk of delivering a low-birth-weight baby at term after an uncomplicated pregnancy varies with maternal circulating concentrations of a placental protein, pregnancy-associated plasma protein-A (PAPP-A) in the first 10 weeks after conception. Poor fetal growth may therefore already have been determined by the time obstetric monitoring begins after completion of the first trimester.

PAPP-A acts as a protease on the binding proteins of insulin-like growth factor (IGF)^{2,3} and may therefore increase the known stimulatory effects of placental IGFs⁴. Circulating concentrations of PAPP-A increase during the first three months of pregnancy, and this protein is highly expressed in trophoblasts⁵. We therefore investigated a possible link between the birth weight of a baby at term and maternal levels of PAPP-A during the first trimester. We determined the specificity of associations with PAPP-A by comparing it with the free β -subunit of human chorionic gonadotrophin (β -CG), another trophoblast-derived protein whose circulating levels change in the first trimester⁶ but which is functionally unrelated to the IGF system.

As part of a prospective, multicentre, non-interventional cohort study, we obtained serum from 4,288 women at 8–12 weeks of gestation (dated by ultrasound, equivalent to 6–10 weeks after conception), who ultimately had uncomplicated singleton pregnancies and delivered normal, live babies at full term. We had complete data for these women on maternal age, parity, height, body-mass index, race and smoking status.

Serum levels of PAPP-A and free β -CG were measured using a Kryptor immunoassay analyser (Brahms Diagnostica) and converted to multiples of the appropriate

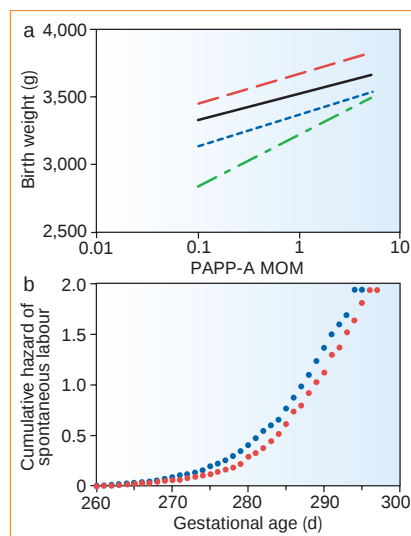


Figure 1 Association between pregnancy-associated plasma protein-A (PAPP-A), birth weight at 38–41 weeks of gestation and timing of labour at full term. **a**, Eventual birth weight plotted against PAPP-A multiples of the median (MOM) on a log₁₀ scale from a linear regression analysis. Curves (from bottom to top) represent gestational ages of 38, 39, 40 and 41 weeks. Coefficients (95% CI) for change in birth weight associated with a one log₁₀ unit change in PAPP-A MOM: 38 weeks, 380 (209–552); 39 weeks, 231 (113–349); 40 weeks, 196 (104–289); 41 weeks, 221 (112–331); $P < 0.0001$ in each case. Coefficients were virtually unchanged after adjusting for age, parity, body-mass index, height, smoking status and race. **b**, Cumulative hazard (Nelson–Aalen cumulative hazard function¹⁰) of spontaneous labour on each day of gestation at full term, comparing the lowest (top curve) and highest (bottom curve) quintiles of first-trimester PAPP-A MOMs. Univariate comparison, $P = 0.0003$ (log rank test).

gestational median (MOM), corrected for smoking status⁷ and maternal weight⁸.

There was a greater proportion of low-birth-weight infants (under 2,500 g) delivered to women with a first-trimester PAPP-A concentration in the lowest 5% (9 out of 201, 4.5%) compared with other women (65 out of 4,087, 1.6%; $P = 0.002$). Using multivariate logistic regression (adjusting for maternal height, race, body-mass index, smoking status and elective delivery), a one log₁₀ unit increase in first trimester PAPP-A MOM (roughly equivalent to the range from the 1st to the 99th percentile) was associated with an 80% reduction in the risk of a low-birth-

weight baby (adjusted odds ratio, 0.2; 95% CI, 0.1–0.6; $P = 0.002$). In contrast, there was no significant independent relationship in the case of free β -CG (adjusted odds ratio, 0.6; 95% CI, 0.2–1.3; $P = 0.17$).

The factors that determine variation in birth weight are fetal growth and the duration of pregnancy. We examined the relationship between PAPP-A concentration and fetal growth using multiple linear-regression analysis, and identified a strong, positive correlation between first-trimester levels of PAPP-A and eventual birth weight at 38–41 weeks of gestation (Fig. 1a). There was no strong association between free β -CG levels and birth weight at the same gestational age.

The relationship between first-trimester concentrations of PAPP-A or free β -CG and the timing of labour at full term was determined using time-to-event analysis⁹. Vaginal delivery after non-induced labour was taken as the event and all other modes of delivery were treated as censored. Lower concentrations of PAPP-A during the first trimester were associated with an earlier onset of spontaneous labour at full term (Fig. 1b).

In a multivariate proportional hazards model, there was a strongly additive and inverse relationship between levels of PAPP-A and free β -CG, and the likelihood of spontaneous labour on any given day of gestation at full term (adjusted hazard ratio for a one log₁₀ unit change in MOM (95% CI): PAPP-A, 0.81 (0.69–0.96), $P = 0.003$; free β -CG, 0.79 (0.69–0.91), $P < 0.001$).

The association between PAPP-A levels, fetal growth and the timing of labour is biologically plausible, as PAPP-A is highly expressed in first-trimester trophoblasts⁵ and may be responsible for activation of IGFs. Our results indicate that the risk of delivering a low-birth-weight baby at full term may be determined by the placental activity of IGFs in very early pregnancy.

Gordon C. S. Smith*, Emily J. Stenhouse†, Jennifer A. Crossley‡, David A. Aitken‡, Alan D. Cameron†, J. Michael Connor‡

*Department of Obstetrics and Gynaecology, University of Cambridge, The Rosie Hospital, Robinson Way, Cambridge CB2 2QQ, UK
e-mail: gcss2@cam.ac.uk

†Department of Fetal Medicine, The Queen Mother's Hospital, Glasgow G3 8SJ, UK

‡Institute of Medical Genetics, Yorkhill NHS Trust, Glasgow G3 8SJ, UK

1. Smith, G. C. S., Smith, M. F. S., McNay, M. B. & Fleming, J. E. *N. Engl. J. Med.* **339**, 1817–1822 (1998).
2. Lawrence, J. B. et al. *Proc. Natl Acad. Sci. USA* **96**, 3149–3153 (1999).
3. Laursen, L. S. et al. *FEBS Lett.* **504**, 36–40 (2001).
4. Rutanen, E. M. *Curr. Opin. Obstet. Gynecol.* **12**, 163–168 (2000).
5. Overgaard, M. T. et al. *Biol. Reprod.* **61**, 1083–1089 (1999).
6. Ozturk, M. et al. *Endocrinology* **120**, 549–558 (1987).
7. Spencer, K. *Prenat. Diagn.* **19**, 1065–1066 (1999).
8. Neveux, L. M., Palomaki, G. E., Larrivee, D. A., Knight, G. J. & Haddow, J. E. *Prenat. Diagn.* **16**, 1115–1119 (1996).
9. Smith, G. C. S. *Hum. Reprod.* **16**, 1497–1500 (2001).
10. Hosmer, D. W. & Lemeshow, S. *Applied Survival Analysis* (Wiley, New York, 1999).

Competing financial interests: declared none.

Sub-laser-cycle electron pulses for probing molecular dynamics

Hiromichi Niikura*, F. Légaré†, R. Hasbani*, A. D. Bandrauk†, Misha Yu. Ivanov*, D. M. Villeneuve* & P. B. Corkum*

* National Research Council of Canada, 100 Sussex Drive, Ottawa, Ontario K1A 0R6, Canada

† Université de Sherbrooke, Sherbrooke PQ, Canada

Experience shows that the ability to make measurements in any new time regime opens new areas of science. Currently, experimental probes for the attosecond time regime (10^{-18} – 10^{-15} s) are being established. The leading approach is the generation of attosecond optical pulses by ionizing atoms with intense laser pulses. This nonlinear process leads to the production of high harmonics during collisions between electrons and the ionized atoms. The underlying mechanism implies control of energetic electrons with attosecond precision. We propose that the electrons themselves can be exploited for ultrafast measurements. We use a ‘molecular clock’, based on a vibrational wave packet in H_2^+ to show that distinct bunches of electrons appear during electron–ion collisions with high current densities, and durations of about 1 femtosecond (10^{-15} s). Furthermore, we use the molecular clock to study the dynamics of non-sequential double ionization.

A substantial effort is under way to develop single attosecond optical pulses^{1–3}, or trains of attosecond pulses^{4–6}, using the physical processes occurring in high-harmonic generation⁷. High harmonics are produced during the electron–ion collisions induced by strong-field laser ionization, usually referred to as ‘recollision’. Within one optical period an electron is removed from the atom, is driven back when the laser field reverses its direction, and collides with the parent ion. The duration of the electron–ion recollision largely determines the duration of the attosecond photon pulse.

Here we study the recollision electron wave packet, measuring both the probability of recollision and its time structure. Although only one electron is involved in the electron–ion recollision, we adopt the language of electron beams to indicate the potential applications of recollision electrons. These applications are the topic of the final section of this Article.

We characterize the unusually large current density and its time structure as seen by the ion following ionization. To do this, we use H_2 molecules in a low-density gas as a molecular clock. As ionization simultaneously forms two wave packets—one a vibrational wave packet moving on the $\text{H}_2^+(X^2\Sigma_g^+)$ surface; the other, the electron wave packet that we wish to study—ionization starts the vibrational clock in H_2^+ . We choose H_2^+ as the molecular clock because of the speed of its vibrational wave packet, and because all excited states of H_2^+ directly dissociate. By choosing the molecular axis perpendicular to the laser electric field, we decouple the $X^2\Sigma_g^+$ and $A^2\Sigma_u^+$ surfaces in H_2^+ , ensuring that the clock remains accurate in the presence of the field. To read the clock, we observe the kinetic energy of the protons produced by inelastic scattering when the electron recollides with the parent ion. The kinetic energy distribution measures the position of the vibrational wave packet at the time of recollision, and therefore the recollision time. In our experiment the time resolution is ~ 1 fs.

Next, we apply the molecular clock to follow the subcycle correlated electron dynamics. Non-sequential double ionization (two-electron ionization that cannot be described by two sequential single-electron ionization processes) is a common occurrence during strong-field ionization of atoms or molecules containing two or more electrons^{7–13}. We distinguish the double ionization due to recollision from instantaneous double ionization by using the molecular clock, and find that electron recollision dominates others by at least two orders of magnitude. We confirm that the most important route to non-sequential ionization is through the production of excited states by recollision that can later ionize in the

strong laser field.

Finally, we compare the double-ionization yield due to recollision in H_2 and helium^{8,11,13}. We find that double ionization (excitation) is about ten times more probable in hydrogen molecules than in helium^{11,13}.

Selecting the fragmentation channel

We now proceed to fully characterize the current density using H_2 double-ionization (excitation) for all aspects of the measurement. (For convenience, we will use ‘double-ionization’ when referring to either the non-sequential emission of two electrons, or the emission of one and the correlated excitation of the other.)

First, we identify collision-induced excitation or double ionization through the previously observed^{14,15} high kinetic energy of the fragment protons that are produced. We show that recollision is responsible for these energetic protons by comparing the kinetic-energy spectrum measured with linear and elliptically polarized light¹⁶. Second, the ellipticity dependence of the proton yield measures the initial velocity spread of the electron wave packet. With this input, we calculate the current density seen by the newly ionized ion. Third, we confirm the predicted temporal structure by comparing the calculated and measured kinetic-energy spectrum. Finally, we confirm the magnitude of the current density by

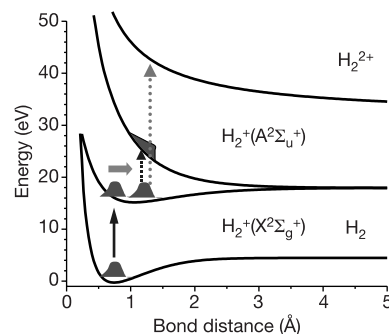


Figure 1 Important potential-energy curves for H_2 and its ions. Ionizing H_2 forms a dual wave packet. One is a vibrational wave packet on the ground (Σ_g^+) state of H_2^+ (shown). The other is an electron wave packet that moves in the laser field. The vibrational wave packet evolves until H_2^+ is excited by inelastic scattering caused by the returning electron wave packet.

comparing the calculated and measured probability of double ionization.

Figure 1 plots the potential-energy surfaces of molecular hydrogen and its ions. Single ionization (represented as the solid vertical arrow in Fig. 1) results in the formation of two correlated wave packets. One is the electron wave packet that we wish to characterize. The other is a vibrational wave packet moving (horizontal arrow) on the field-modified Σ_g potential-energy surface until electron re-collision occurs.

When the electron wave packet returns to the ion it can inelastically scatter, producing excited states of H_2^+ or further ionizing the ion (dotted arrow in Fig. 1). All excited states of H_2^+ lead to dissociation of the molecular ion. They can be identified through the kinetic-energy spectrum of the protons, provided that the laser field does not further ionize them. To avoid further ionization we keep the light intensity low, and we exploit two additional effects: that ionization is not enhanced for perpendicular orientation¹⁷, and that ionization is suppressed for anti-symmetric states¹⁸. The kinetic energy spectrum gives us a relatively uncomplicated 'time history' of the recollision dynamics. The channel of interest produces fragments with kinetic energy in the range 1–10 eV.

We performed the experiment in a time-of-flight mass spectrometer containing H_2 molecules at a pressure of 10^{-6} torr. We apply a uniform electric field across two parallel electrodes separated by 3 cm. A 1-mm-diameter hole in one electrode allows us to observe protons only from those molecules whose axis lies close to parallel with respect to the time-of-flight axis. The signal is measured on a microchannel plate detector.

Although we could observe molecules with any orientation, we concentrate on those that are oriented perpendicular to the laser polarization. H_2^+ cannot change its orientation¹⁹ in the brief interval between ionization, recollision and dissociation. The collection half-angle depends on the dissociation velocity of the fragments and the applied electric field. For our experimental conditions, the half angle was 8° at 10 eV with the angle increasing with the inverse of the velocity.

H_2 molecules are ionized by a 50-fs light pulse of 800 nm wavelength, propagating perpendicular to the time-of-flight axis and focused to a peak laser intensity of $(1.5 \pm 0.5) \times 10^{14} \text{ W cm}^{-2}$. The intensity is calibrated against the ionization of xenon²⁰. At this intensity the recollision electron should have maximum collision energy of ~ 30 eV, sufficient only to excite $X^2\Sigma_g^+ \rightarrow A^2\Sigma_u^+$ in the molecular ion. To change the ellipticity of the laser light we rotate a half-wave plate that is placed before a fixed quarter-wave plate. This procedure preserves the direction of the major axis as the ellipticity

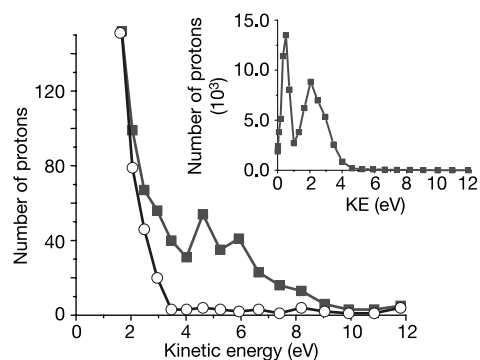


Figure 2 The number of protons measured per unit energy as a function of the proton kinetic energy. The upper curve is obtained with linear polarized light, whereas the lower curve is obtained with elliptically polarized light (ellipticity $E_y/E_x = 0.3$; E_x and E_y are defined in the text). Intermediate ellipticities fall between these curves. The main plot is for perpendicularly aligned molecules, the inset is for molecules aligned parallel to the laser field (KE, kinetic energy).

changes. We now proceed to identify the signature of recollision in the kinetic-energy spectrum.

Confirming recollision

Figure 2 plots the measured kinetic energy spectrum of the protons. The main plot, shown on a very-much-expanded vertical scale, is for molecules oriented perpendicular to the field direction, whereas the inset is for molecules oriented in the parallel direction. Because of its scale, the main plot of Fig. 2 most clearly shows the low-probability events, namely the protons with high kinetic energy.

To identify the channel of interest, we must distinguish it from other dissociation channels that influence the kinetic-energy spectrum of the protons. Bond softening^{21,22}, which originates from the mixing of the Σ_g and Σ_u levels by the laser field, is responsible for the kinetic-energy peak near 0.5 eV, clearly seen in Fig. 2 inset. To eliminate bond-softened molecules, we select perpendicularly oriented molecules where Σ_g and Σ_u are decoupled. Enhanced ionization, which refers to the large and broad increase in ionization rate with the internuclear coordinate in the region where the bond breaks^{17,23,24}, is responsible for the peak in the ~ 2.5 eV region. Like bond softening, it is also eliminated for perpendicular molecules¹⁸.

We confirm, by measuring the number of protons as a function of the ellipticity of the laser light^{1,10,12,16}, that recollision is responsible for the high-kinetic-energy protons. Processes caused by electron recollision are very sensitive to small deviations from linear polarization⁷. The inset in Fig. 3 illustrates the method that we use. In linearly polarized light, the laser field dominates the electron motion in the direction of the laser polarization. However, the quantum-mechanical uncertainty of the electron's velocity at the time of ionization causes the electron wave packet to have expanded significantly in the lateral direction by the time it re-collides with the ion (Fig. 3 inset). We can easily influence the electron trajectory by adding a small ellipticity to the laser pulse. The weak perpendicular field pushes the electron wave packet to the side so that recollision becomes impossible.

The sensitivity to ellipticity is seen in Fig. 2. The upper curve is obtained using linearly polarized light, and the lower curve is obtained using light with ellipticity of $E_y/E_x = 0.3$ (where E_y (E_x) is the laser electric field in the direction perpendicular (parallel) to the molecular axis). Ellipticities between 0 and 0.3 occupy the intermediate region, whereas the lower curve is essentially unchanged for higher ellipticities. The difference of data counts between the upper and lower curves determines the kinetic-energy

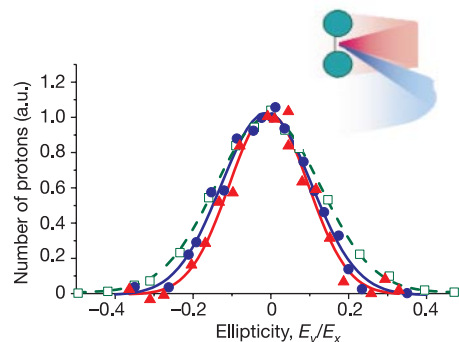


Figure 3 Ellipticity dependence of the number of energetic protons produced when an intense laser field ionizes H_2 . Such narrow ellipticity dependence is characteristic of inelastic scattering between the recollision electron and the ion. The data is for H_2 aligned both parallel (blue filled circles) and perpendicular (red filled triangles) to the laser field. For comparison purposes, the ellipticity dependence of the double-ionization probability of argon, an atom with almost the same ionization potential, is shown (open green squares). The inset illustrates that elliptically polarized light can redirect the electron so that recollision becomes impossible (see text). (a.u., arbitrary units.)

spectrum of protons produced by recollision, as we will show below. It is clear from Fig. 2 that our measurements should concentrate on kinetic energies above about 4 eV.

Figure 3 shows the ellipticity dependence of the recollision yield. All high-kinetic-energy fragments have the same ellipticity dependence. At each ellipticity, we integrate the signal count in the kinetic energy range 4–9 eV. The three curves included in Fig. 3 are for argon and for H₂ molecules aligned parallel and perpendicular to the laser polarization.

Lateral spread of the electron wave packet

The velocity that the electron acquires in the direction perpendicular to the laser field during ionization determines the expansion of the electron wave packet in the lateral direction. If the velocity is large, the wave packet that recollides is spread over a large area. We concentrate on the lateral velocity for molecules aligned perpendicular to the laser field.

As the laser ellipticity increases, the electron wave packet is pushed further from the ion core in the direction of the minor component of the laser field. When the electron wave packet returns to the ion core for the first time, at a time $\Delta t = 1.8$ fs after its birth, its displacement is given by $dx \approx 5.14\varepsilon E/m\omega^2$ (ref. 12), where ε is ellipticity, E is the strength of laser field, m is the electron mass and ω is the laser angular frequency. Our $1/e$ width $\varepsilon = 0.14$ yields a $1/e$ displacement $dx = 7.7$ Å. Because the effective collision cross-section is small (~ 1 Å²), the displacement allows us to measure the spatial distribution of the electron wave packet when it returns to the ion core¹². From this distribution, we determine that the initial velocity spread at the time of ionization is ~ 4.2 Å fs⁻¹. The observed ellipticity dependence shows a gaussian form, as predicted by tunnelling theory²⁵. Argon is used as a reference that we can accurately calculate. For argon (open squares in Fig. 3) the same procedure yields ~ 5.6 Å fs⁻¹, in excellent agreement with 5.4 Å fs⁻¹ predicted by the atomic tunnelling theory²⁵.

Knowledge of both the expansion velocity of the electron wave packet and its shape provides the initial conditions for a semi-classical simulation of the electron dynamics following ionization¹³. In the calculation, we follow many electron trajectories in two dimensions; each trajectory is characterized by different initial positions and velocities and by a weight ρ_i , which reflects the probability of that trajectory. We count the number of trajectories passing through the surface of a circle of radius r_0 , where πr_0^2 is the inelastic cross-section (ref. 26 and <http://physics.nist.gov/PhysRefData/Ionization/Xsection.html>) for the $\Sigma_g \rightarrow \Sigma_u$ transition in H₂⁺, within a unit time. In the calculations, we correct the atomic tunnelling²⁷ rate to include the molecular structure. We also include the influence of the molecular ion's field on the free electron trajectories.

Figure 4 is a plot of the calculated current density experienced by the molecular ion. Only electrons with energy greater than the

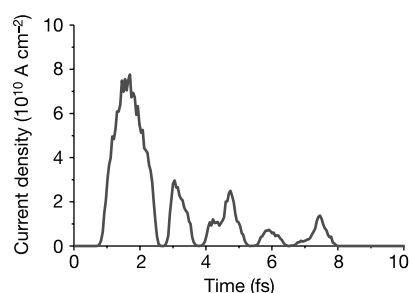


Figure 4 The calculated electron current density experienced by the ion as a function of time after ionization. Only electrons with sufficient energy for inelastic scattering are included.

resonant energy between Σ_g and Σ_u are included in this plot. Otherwise, the current density includes all other electrons that pass through the collisional surface of area πr_0^2 . Figure 4 shows that the recollision current consists of five micro-bunches. The current density rises in the second half-period after ionization, reaching 8×10^{10} A cm⁻². At later times, the amplitude decreases rapidly. Such high current densities are available only from very large accelerators such as SLAC.

Temporal structure of the electron wave packet

To confirm the time structure of the recollision, we turn our attention to the vibrational wave packet. The laser parameters place our experiment in the regime where ionization can be calculated by assuming that the electron tunnels through the barrier caused by the superposition of the laser field and the ion field^{27,28}. Removing an electron from H₂ forms a vibrational wave packet on the ground Σ_g of H₂⁺ that resembles the ground-state wavefunction. Ionization simultaneously releases the electron wave packet that we have been discussing. The distortion of the vibrational wave packet compared to the ground-state wavefunction of H₂ (ref. 29) owing to the dependence of the ionization rate on the internuclear separation is included in the calculation.

We follow the H₂⁺ wave packet by solving Schrödinger's equation on the Σ_g molecular potential. By including the interaction of the laser field with the perpendicular induced dipole, we confirm that, at this intensity, other states have a negligible effect on the motion of the wave packet. The proton kinetic-energy spectrum resulting from recollision is obtained by projecting the Σ_g wave packet onto the Σ_u continuum wavefunctions.

Figure 5 (solid line) shows the incoherent sum of the kinetic-energy spectra of protons produced by all five micro-bunches weighted according to their relative probabilities as determined by the calculated time-independent current density in Fig. 4, and weighted by the cross-section at the average internuclear separation²⁶ for each micro-bunch. We have used angle-averaged cross-sections because, to our knowledge, no cross-section data are available for oriented molecules. We show the individual contribution of the first and third recollision as dotted lines. Their time separation is 2.7 fs. It is clear from Fig. 5 that we can resolve contributions from an intermediate electron pulse.

The experimental kinetic energy spectrum (data points) is also plotted in Fig. 5. In the vertical direction, error bars are determined by shot-to-shot statistics. The error in the horizontal direction is only dependent on the measurement resolution (500 ps) of the time-of-flight mass spectrometer. The error is of the order of the spacing of the data points. To obtain the experimental data for Fig. 5, the kinetic-energy spectrum, measured with elliptically polarized light (Fig. 2, circles), was subtracted from the kinetic-energy spectrum measured with linearly polarized light (Fig. 2, squares).

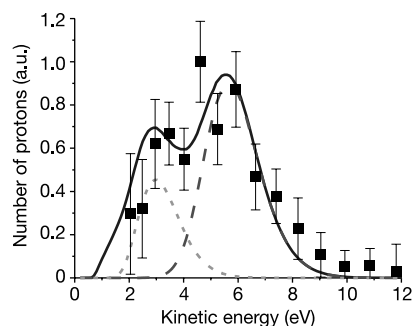


Figure 5 Observed (data points) and calculated (solid curve) kinetic-energy spectrum for the energetic protons caused by inelastic scattering. The agreement confirms the time dependence of the current density. The dashed lines show the contributions from the first (long dash) and third electron micro-bunch (short dash).

This procedure distinguishes the processes due to recollision from other sources.

The overlap between the calculation and the experiment in Fig. 5 is notable. We stress that our ability to measure the ~ 1 -fs duration of the first micro-bunch is achieved because of the double correlation between the electronic and vibrational wave packets. We do not use either time-dependent polarization¹ or a few-cycle pulse^{2,3}.

Magnitude of the electron current density

In Fig. 5, we account for the relative magnitude of each recollision peak. The only free parameter is the overall normalization of the experimental kinetic-energy spectrum with respect to the theoretical one. Theoretically, using the current density in Fig. 4 and the angle-averaged cross-section, we can predict the angle-averaged probability that one single-ionization event will lead to inelastic excitation of Σ_u . At $1.5 \times 10^{14} \text{ W cm}^{-2}$, we calculate that 7% of all ionization events result in inelastic scattering. This establishes the vertical scale in Fig. 5 for the calculations. We now establish this experimentally.

The number of protons produced by recollision is plotted in Fig. 6 as a function of the alignment of the molecular axis with respect to the laser field. There are about 10 times more inelastic scattering events for molecules aligned parallel to the field than perpendicular. Were we able to align all molecules relative to the laser polarization, we could separate the influence of angle-dependent ionization rates and cross-sections. But owing to the small asymmetry in the polarizability and the large rotational constant, alignment is not possible.

Figure 7 shows the branching ratio between H_2^+ (circles), bond softening ($0\text{--}0.9 \text{ eV}$, upward pointing triangles), enhanced ionization ($0.9\text{--}4 \text{ eV}$, downward pointing triangles) and inelastic scattering ($>4 \text{ eV}$, squares). To obtain this branching ratio, we measure the probability of each product channel at 10° intervals. This establishes the probability as a function of angle. Then we integrate each channel over all 4π steradians. The branching ratio of each channel is obtained by dividing each probability by the sum of all channels. The experimental branching ratio into inelastically scattered fragments is $\sim 2\text{--}3\%$ at $1.5 \times 10^{14} \text{ W cm}^{-2}$. This establishes the vertical scale in Fig. 5 for the measurement. As with helium, double ionization is almost independent of the laser intensity^{11,13}.

Both the calculation and the measurement have a degree of uncertainty. By assigning a fragment kinetic energy $>4 \text{ eV}$ to inelastic scattering, we experimentally underestimate the true branching ratio into inelastic channels. We know that, at least for perpendicular molecules (Fig. 2), about 30% of all inelastic events lead to protons with kinetic energy in the $2\text{--}4 \text{ eV}$ range. But if the molecule is oriented parallel to the internuclear axis, such events are

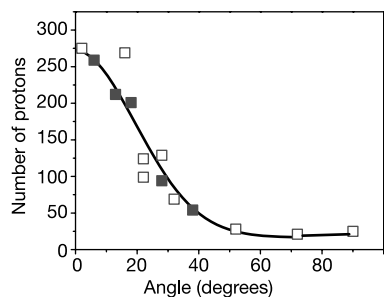


Figure 6 Angular dependence of the double-ionization (excitation) probability due to recollision. Data were taken from 40° on one side of parallel to 90° on the other side to ensure that the results were symmetric. All data points are included in the figure. Filled data points represent data on one side (0 to -40°) and the open points represent the other side (0 to 90°). Inelastic scattering is about 10 times more likely for a molecule aligned parallel to the field than for a molecule aligned perpendicular to the field.

overwhelmed by the enhanced ionization peak and we cannot accurately measure them. Theoretically, we overestimate the experimental branching ratio by using hard-sphere¹³, angle-averaged cross-sections and two-dimensional trajectories. In view of these uncertainties, the experiment and calculation agree.

Strong-field non-sequential double ionization

Of the phenomena observed in the interaction between atoms/molecules and ultrashort intense laser pulses (high-harmonic generation^{30,31}, very-high-energy electron production^{32,33}, and non-sequential double ionization^{8,10,11}), all but the last can be described with a single active electron¹¹. Thus non-sequential double ionization gains a particular significance within strong-field physics, connecting it to multi-electron physics, a subject that is important in many other areas of science. There are a number of important issues in non-sequential double-ionization physics that our experiment allows us to comment on.

Multi-electron dynamics. Although we have expressed our results in terms of the current density and concentrated on the electron that recollides, we obtain the current by measuring the correlated-electron dynamics during strong-field ionization. Until recently, the basic mechanism responsible for double ionization was a subject of controversy¹¹. Strong-field double ionization was initially thought to be caused by shake-off—the departing electron causes the remaining electron to become unbound. It is now accepted that recollision is the main mechanism responsible for double ionization. However, previous experimental measurements were not precise enough to eliminate the possibility that shake-off (or other instantaneous double-ionization processes that are important for X-ray double ionization) plays a minor role.

The double correlation of vibrational and electron wave packet in H_2 allows instantaneous double ionization to be identified, as it would place the H_2^+ on an excited surface before any nuclear motion could occur in the Σ_g . Referring to Fig. 2, we do not observe protons with kinetic energy above $\sim 9.5 \text{ eV}$. The lack of energetic protons with kinetic energy greater than 10 eV implies that, at an intensity of $1.5 \times 10^{14} \text{ W cm}^{-2}$, the branching ratio into instantaneous double ionization cannot exceed 10^{-4} .

However, we do resolve in time the decay of the electron–electron interaction in H_2 during ionization in a strong field. Following ionization, the interaction has the complex temporal structure similar to the electron current density in Fig. 4.

The role of excited states. The unique advantage of H_2 is that all excited states of H_2^+ are unstable. Therefore the kinetic-energy spectrum of the fragments can identify the dissociation channels. We find that excited states dominate H_2 double ionization in the intensity regime in which we work. At higher intensities, the Σ_u state would ionize during the laser pulse, making it an important pathway to double ionization. The role of excitation as a route to double

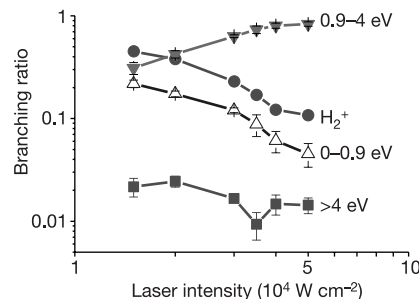


Figure 7 Intensity dependence of the double-ionization (excitation) probability. As with helium, the double-ionization probability (squares) is almost independent of the laser intensity in the non-sequential range. The other channels are bond softening (upward-pointing triangles), enhanced ionization (downward pointing triangles) and H_2^+ (circles).

ionization in atoms has been the subject of considerable speculation^{13,34,35}.

The significance of alignment in molecular ionization. The transverse velocity spread of the electron as it leaves the atom (molecule) determines the probability of all recollision processes (high-harmonic generation, double ionization and elastic scattering). The data in Fig. 3 allow us to make three important observations about the transverse velocity spread. First, for either alignment of the molecule with respect to the laser polarization, the ellipticity dependence shown in Fig. 2 is independent of the kinetic energy of the fragments in the range caused by recollision (4–9 eV), and is therefore independent of which micro-pulse leads to the excitation. Second, the ellipticity dependence is slightly different for the parallel and perpendicular orientation of the molecules. The transverse velocity spread perpendicular to the laser field therefore depends slightly on the molecular alignment with respect to the laser field. For molecules aligned parallel to the laser polarization, the 1/e width of the velocity distribution is 5.0 Å fs^{-1} , while, for perpendicular aligned molecules, the 1/e width of the velocity distribution is 4.2 Å fs^{-1} . Third, although Ar has almost the same ionization potential as H_2 , the 1/e width of the transverse velocity of Ar ($\sim 5.6 \text{ Å fs}^{-1}$) is larger than that for either orientation of H_2 .

In contrast to atomic tunnelling theory²⁵, there have been no theoretical predictions of the electron wave packet spread in molecules. Qualitatively, in agreement with our measurements, we might expect a smaller spread from the broader tunnel that characterizes a perpendicular molecule.

The probability of double ionization (excitation). Although both H_2 and helium have two electrons, the probability of double ionization or excitation is about 10 times higher for H_2 than for He (ref. 11). It is also significantly greater for H_2 than for neon³⁶, although neon has many more electrons.

Using attosecond electrons for probing

Implicit in our results is the potential to exploit recollision electrons in a number of subfields of gas-phase molecular science operating on all timescales. Currently it is assumed that attosecond science will be an extension of ultrashort-pulse science in a new time regime. However, extreme ultraviolet attosecond pulses are not generated in a laser system but in a very inefficient process involving ‘free’ electrons^{2,3}. Focusing the attosecond extreme ultraviolet radiation onto another gas is also inefficient. In contrast, attosecond electron bunches are produced with $\sim 100\%$ efficiency, and have a high probability of interaction. It seems likely that attosecond science will use attosecond electron pulses and attosecond photon pulses equally. In addition, the technology needed for producing and using attosecond electron bunches is available in many laboratories. We have used a 50-fs laser to produce an electron micro-bunch that has a duration of ~ 1 fs. This is possible because we exploit correlation—we probe only those molecules that undergo ionization.

If it is necessary to eliminate the weaker micro-bunches and reach individual bunches of electrons, we can borrow two approaches that have been proposed to reach single attosecond optical pulses. One approach uses a few-cycle visible pulse to confine the time of ionization, and to ensure that the driving pulse is terminated before a second recollision is likely^{2,3}. The other approach exploits a fundamental pulse with time-dependent polarization to control the electron trajectory¹. With time-dependent polarization, recollision is only possible for a fraction of the driving laser period, resulting in attosecond extreme ultraviolet pulses and attosecond recollisions.

But what are the implications for conventional femtosecond science? Conventional femtosecond science has no convenient short-wavelength source, although there is a great deal of interest in creating one using X-rays³⁷ or electrons³⁸. Recollision electrons can play that role. Depending on the wavelength of the laser light

that creates them, recollision electrons can have very high energies. These electrons are produced when and where they are needed. In effect, they are ‘slaved’ with attosecond precision to the optical beam that created them. Controlling the laser pulse controls the electrons, both their production and their subsequent motion^{39,40}.

As probes in femtosecond experiments, recollision electrons are expected to have unique advantages over laser photons. The most important advantage arises during elastic scattering. If the electrons are sufficiently energetic, an image of the structure of the molecule is impressed onto the diffracted electron distribution. Although the recollision electron probes the ion a few femtoseconds after ionization (for most molecules), there is no time for the structure of the neutral molecule to change. Therefore, although the ion is probed, the structure of the neutral molecule is imaged.

Even when time resolution is not an issue, recollision electrons will find applications. For example, in a gas that combines a mixture of excited and unexcited molecules, the intense laser field that is used to create the recollision electrons will preferentially ionize the excited molecules—exactly the molecules that we wish to observe. As the recollision electron probes only its parent, excited molecules will be preferentially imaged.

Few molecules will be needed for imaging, because the current densities are so large. In fact, there is a substantial effort under way to develop methods for single-molecule imaging or few-molecule imaging⁴⁰. The very large current densities that characterize recollision events suggest that recollision electrons could contribute to this effort. □

Received 20 December 2001; accepted 30 April 2002; doi:10.1038/nature00787.

- Corkum, P. B., Ivanov, M. Yu. & Burnett, N. H. Sub-femtosecond pulses. *Opt. Lett.* **19**, 1870–1872 (1994).
- Drescher, M. *et al.* X-ray pulses approaching the attosecond frontier. *Science* **291**, 1923–1927 (2001).
- Hentschel, M. *et al.* Attosecond metrology. *Nature* **414**, 509–513 (2001).
- Antoine, P., L’Huillier, A. & Lewenstein, M. Attosecond pulse trains using high-order harmonics. *Phys. Rev. Lett.* **77**, 1234–1237 (1996).
- Paul, P. M. *et al.* Observation of a train of attosecond pulses from high harmonic generation. *Science* **292**, 1689–1692 (2001).
- Papadogiannis, N. A., Witzel, B., Kalpouzou, C. & Charalambidis, D. Observation of attosecond light localization in higher order harmonic generation. *Phys. Rev. Lett.* **83**, 4289–4292 (1999).
- Corkum, P. B. A plasma perspective on strong field multiphoton ionization. *Phys. Rev. Lett.* **71**, 1994–1997 (1993).
- Fittinghoff, D. N., Bolton, P. R., Chang, B. & Kulander, K. C. Observation of nonsequential double ionization of helium with optical tunneling. *Phys. Rev. Lett.* **69**, 2642–2645 (1992).
- Weber, Th. *et al.* Correlated electron emission in multiphoton double ionization. *Nature* **405**, 658–661 (2000).
- Bhardwaj, V. R., Rayner, D. M., Villeneuve, D. M. & Corkum, P. B. Quantum interference effects in double ionization and fragmentation of C_6H_6 . *Phys. Rev. Lett.* **87**, 253003–1–253003–4 (2001).
- Walker, B. *et al.* Precision measurement of strong field double ionization of helium. *Phys. Rev. Lett.* **73**, 1227–1230 (1994).
- Dietrich, P., Burnett, N. H., Ivanov, M. Yu. & Corkum, P. B. High harmonic generation and correlated two electron multiphoton ionization with elliptically polarized light. *Phys. Rev. A* **50**, 3585–3588 (1994).
- Yudin, G. L. & Ivanov, M. Yu. Physics of correlated double ionization of atoms in intense laser fields: Quasistatic tunneling limit. *Phys. Rev. A* **63**, 033404–1–033404–14 (2001); erratum *Phys. Rev. A* **64**, 019902 (2001).
- Trump, C., Rottke, H. & Sandner, W. Strong-field photoionization of vibrational ground-state H_2^+ and D_2^+ molecules. *Phys. Rev. A* **60**, 3924–3928 (1999).
- Staudte, A. *et al.* Observation of a nearly isotropic, high energy Coulomb explosion group in the fragmentation of D_2 by short laser pulses. *Phys. Rev. A* **65**, 020703–1–020703–4 (2002).
- Sakai, H. *et al.* Non-sequential double ionization of D_2 molecules with intense 20 fs pulses. *Phys. Rev. Lett.* (submitted).
- Constant, E., Stapelfeldt, H. & Corkum, P. B. Observation of enhanced ionization of molecular ions in intense laser fields. *Phys. Rev. Lett.* **76**, 4140–4143 (1996).
- Muth-Böhm, J., Becker, A. & Faisal, F. H. M. Suppressed molecular ionization for a class of diatomics in intense femtosecond laser fields. *Phys. Rev. Lett.* **85**, 2280–2283 (2000).
- Ellert, Ch. & Corkum, P. B. Disentangling molecular alignment and enhanced ionization in intense laser fields. *Phys. Rev. A* **59**, R3170–R3173 (1999).
- Hankin, S., Villeneuve, D. M., Corkum, P. B. & Rayner, D. M. Nonlinear ionization of organic molecules in high intensity laser fields. *Phys. Rev. Lett.* **84**, 5082–5085 (2000).
- Zavriyev, A., Bucksbaum, P. H., Muller, H. G. & Schumacher, D. W. Ionization and dissociation in H_2 in intense laser fields at 1.064 μm , 532 nm and 355 nm. *Phys. Rev. A* **42**, R5500–R5513 (1990).
- Zavriyev, A., Bucksbaum, P. H., Squier, J. & Salin, F. Light-induced vibrational states in H_2^+ and D_2^+ in intense laser fields. *Phys. Rev. Lett.* **70**, 1077–1080 (1993).
- Seideman, T., Ivanov, M. Yu. & Corkum, P. B. The role of electron localization in intense field molecular ionization. *Phys. Rev. Lett.* **75**, 2819–2822 (1995).
- Zuo, T. & Bandrauk, A. D. Charge-resonance-enhanced ionization of diatomic molecular ions by intense lasers. *Phys. Rev. A* **52**, R2511–R2514 (1995).

25. Delone, N. B. & Krainov, V. P. Energy and angular electron spectra for the tunnel ionization of atoms by strong low-frequency radiation. *J. Opt. Soc. Am. B* **8**, 1207–1212 (1991).
26. Peek, J. M. Inelastic scattering of electrons by the hydrogen molecular ion. *Phys. Rev.* **134**, A877–A883 (1964).
27. Ammosov, M. V., Delone, N. B. & Krainov, V. P. Tunnel ionization of complex atoms and of atomic ions in an alternating electromagnetic field. *Zh. Eksp. Teor. Fiz.* **91**, 2008–2013 (1986); *Sov. Phys. JETP* **64**, 1191–1194 (1989).
28. Keldysh, L. V. Ionization in the field of a strong electromagnetic wave. *J. Exp. Theor. Phys.* **47**, 1945–1957 (1964).
29. Dietrich, P., Ivanov, M. Yu., Ilkov, F. & Corkum, P. B. Two-electron dissociative ionization of H₂ and D₂ in infrared fields. *Phys. Rev. Lett.* **77**, 4150–4153 (1996).
30. Macklin, J. J., Kmetz, J. D. & Gordon, C. L. High-order harmonic generation using intense femtosecond pulses. *Phys. Rev. Lett.* **70**, 766–769 (1993).
31. Bartels, R. *et al.* Shaped-pulse optimization of coherent emission of high-harmonic soft X-rays. *Nature* **406**, 164–166 (2000).
32. Mohideen, U. *et al.* High intensity above-threshold ionization of He. *Phys. Rev. Lett.* **71**, 509–512 (1993).
33. Paulus, G. G., Nicklich, W., Xu, H., Lambropoulos, P. & Walther, H. Plateau in above threshold ionization spectra. *Phys. Rev. Lett.* **72**, 2851–2854 (1994).
34. Feuerstein, B. *et al.* Separation of recollision mechanisms in nonsequential strong field double ionization of Ar: The role of excitation tunneling. *Phys. Rev. Lett.* **87**, 043003–043007 (2001).
35. Becker, A. *et al.* Laser-induced inner shell vacancies in double ionized argon. *J. Phys. B* **33**, L547–L552 (2000).
36. Bhardwaj, V. R. *et al.* Few cycle dynamics of multiphoton double-ionization. *Phys. Rev. Lett.* **86**, 3522–3525 (2001).
37. Rischel, C. *et al.* Femtosecond time-resolved X-ray diffraction from laser-heated organic films. *Nature* **390**, 490–492 (1997).
38. Ihee, H. *et al.* Direct imaging of transient molecular structures with ultrafast diffraction. *Science* **291**, 458–462 (2001).
39. Ivanov, M. Yu., Corkum, P. B., Zuo, T. & Bandrauk, A. D. Routes to control of intense-field atomic polarizability. *Phys. Rev. Lett.* **74**, 2933–2937 (1995).
40. Neutze, R. *et al.* Potential for biomolecular imaging with femtosecond X-ray pulses. *Nature* **406**, 752–757 (2000).

Acknowledgements

F.L. acknowledges financial support from Canada's Natural Science and Engineering Research Council, the Canadian Institute for Photonics Innovation and Quebec's Fonds pour la Formation des Chercheurs et l'Aide à la Recherche.

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to P.B.C. (e-mail: Paul.Corkum@nrc.ca).

Magnification of light from many distant quasars by gravitational lenses

J. Stuart B. Wyithe & Abraham Loeb

Astronomy Department, Harvard University, 60 Garden Street, Cambridge, Massachusetts 02138, USA

Exceptionally bright quasars with redshifts up to $z = 6.28$ have recently been discovered¹. Quasars are thought to be powered by the accretion of gas onto supermassive black holes at the centres of galaxies. Their maximum (Eddington) luminosity depends on the mass of the black hole, and the brighter quasars are inferred to have black holes with masses of more than a few billion solar masses. The existence of such massive black holes poses a challenge to models for the formation of structures in the early Universe^{2,3}, as it requires their formation within one billion years of the Big Bang. Here we show that up to one-third of known quasars with $z \approx 6$ will have had their observed flux magnified by a factor of ten or more, as a consequence of gravitational lensing by galaxies along the line of sight. The inferred abundance of quasar host galaxies, as well as the luminosity density provided by the quasars, has therefore been substantially overestimated.

Gravitational lensing leads to a magnification by a factor μ of the apparent source luminosity. Let us first consider the implications of lensing for the early existence of massive quasar systems within structure formation theory. The four highest redshift quasars known¹ (with $z \geq 5.8$; SDSS 1044-0125 was later found⁴ to have $z = 5.73$), were selected in the SDSS (Sloan Digital Sky Survey) photometric system to have magnitudes $z^* < 20.2$ and colours $i^* - z^* > 2.2$. The masses of the central black holes powering these quasars are estimated to be $3 \times 10^9 M_\odot$, where $M_\odot$ is the solar mass, implying³ that the mass of their host galaxies is $\geq 10^{13} M_\odot$. The evolution of the space density of halos of a given mass is described by the Press–Schechter⁵ mass function. Such massive hosts lie on the steep exponential tail of this mass function, so that a correction to the inferred black-hole mass severely affects the estimated cosmological density of galaxy halos which are sufficiently massive to host the observed quasars.

Using the Eddington luminosity to set a lower limit on the inferred black hole mass³, we find that the magnification due to lensing lowers the minimum black hole mass by a factor of μ . This in turn implies that the black hole can form in a lower-mass galaxy. Figure 1 shows the resulting enhancement factor in the space density of galaxy halos that could host the observed SDSS quasars at $z = 6$. The inclusion of lensing has a very large effect (by orders of magnitude) on the expected abundance of such hosts at early cosmic times. The prominence and duty cycle of such hosts has important implications^{6,7} with respect to whether the cosmic neutral hydrogen, a cold remnant from the Big Bang, was re-ionized by star light or by quasars^{8,9}.

For the lensing calculation we must specify the quasar luminosity function (number per co-moving volume per unit luminosity). At $z \lesssim 3$ this is well described by a double power law¹⁰, with all redshift dependence in the evolution⁸ of the break luminosity $L_\star$:

$$L_\star(z) = L_\star(0)(1+z)^{-(1+\alpha_q)} e^{\xi z} \frac{1 + e^{\xi z}}{e^{\xi z} + e^{\zeta z}} \quad (1)$$

We find that an intrinsic luminosity function having slopes at the faint and bright ends of $\beta_l = 1.64$ and $\beta_h = 3.43$, and parameters $L_\star(0) = 1.5 \times 10^{11} L_\odot$, where $L_\odot$ is the solar luminosity, $\alpha_q = -0.5$, $z_\star = 1.45$, $\xi = 2.9$ and $\zeta = 2.7$ adequately (after consideration of the gravitational lensing described below) describes the luminosity

function at $z \lesssim 3$, and the number density of quasars with absolute B -magnitude $M_B < -26$ at $z \approx 4.3$ (measured by SDSS¹¹) and $M_B < -27.6$ at $z \approx 6.0$. The parameter α_q is the slope assumed for the typical quasar continuum $L(\nu) \propto L^{\alpha_q}$. We use the integral luminosity function $N(> L_{\text{lim}}, z) = \int_{L_{\text{lim}}}^{\infty} dL \phi(L, z)$, where L_{lim} is the luminosity of a quasar at redshift z corresponding to an apparent magnitude $z_{\text{lim}}^* L_{\text{lim}}$ was determined from $z_{\text{lim}}^* = 20.2$ using a luminosity distance and a k -correction computed from a model quasar spectrum including the mean absorption by the intergalactic medium¹².

Gravitational lensing is expected to be highly probable for very luminous quasars¹³. Consider a fictitious gravitational lens that always produces a magnification of $\mu = 4$ for the sum of multiple images (the average value for a singular isothermal sphere, SIS) but $\mu = 1$ otherwise. We define τ_{mult} as the probability that a random quasar selected in the source plane will be multiply imaged¹⁴, and F_{MI} to be the magnification-biased probability that an observed quasar will be multiply imaged. Surveys for quasars at $z < 3$ have limiting magnitudes fainter than the break magnitude ($m_B \approx 19$). Whereas $\tau_{\text{mult}} \approx 0.002$ at $z \approx 2$, a survey for quasars to a limit L_{lim} will find a number of lensed sources that is larger by a bias factor of $N(< L_{\text{lim}}/\mu, z)/N(< L_{\text{lim}}, z)$. At $z = 2$, $L_{\text{lim}}(z) \ll L_\star(z)$ and for $\beta_l = 1.6$ the magnification bias factor is about 2.3, resulting in $F_{\text{MI}} \approx 0.005$. At $z = 6$, $\tau_{\text{mult}} \approx 0.008$ is significantly higher¹⁵ than at $z = 2$. Furthermore, the limiting magnitude of the $z \approx 5.8$ survey is significantly brighter than the break magnitude and so $\beta_h = 3.4$ and the bias factor rises to approximately 28. Under these circumstances, F_{MI} is about 0.22. These simple arguments are consistent with previous estimates¹⁶ and demonstrate that lensing has a strong effect on observations of the bright SDSS quasars at $z \geq 5.8$.

To find the magnification bias more accurately, we have computed the probability distributions $dP_{\text{sing}}/d\mu$ and $dP_{\text{mult}}/d\mu_{\text{tot}}$ for the magnification μ of randomly positioned singly imaged sources, and for the sum of magnifications μ_{tot} of randomly positioned multiply imaged sources due to gravitational lensing by foreground

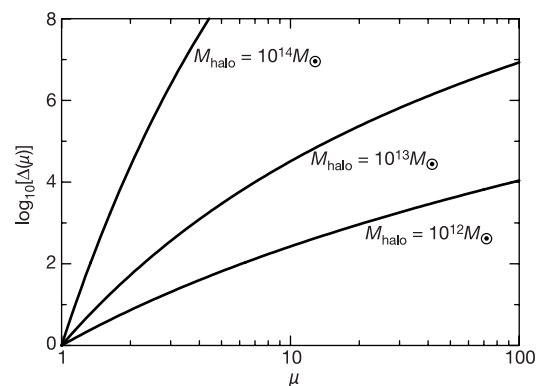


Figure 1 Enhancement in the co-moving space density of host galaxies for quasars at $z = 6$ as a function of the magnification due to lensing, μ . Ignoring lensing, the Eddington limit implies³ a minimum black hole mass of about $3 \times 10^9 M_\odot$ for the SDSS quasars at $z \approx 6$. The inclusion of lensing reduces the implied (minimum) black hole mass by a factor of μ . Given some fixed efficiency for assembling gas into a central black hole within a galaxy, the implied mass of the host galaxy is also lowered by μ after lensing is included. Plotted is $\Delta(\mu) = \{dn/d(\log[M/\mu])\}/\{dn/d(\log[M])\}$ where $dn/d(\log[M])$ is the Press–Schechter⁵ co-moving density of galaxy halos with mass M per logarithmic interval in M . We show curves for three possible halo masses of quasar hosts, $M_{\text{halo}} = 10^{12} M_\odot$, $10^{13} M_\odot$, and $10^{14} M_\odot$. As seen, the enhancement in the implied abundance of quasar hosts increases with increasing halo mass. The above three choices for halo masses are all conservatively lower than inferred for the same black hole mass in the local universe³⁰. Throughout this work we assume the standard cosmological parameters $\Omega_m = 0.35$, $\Omega_\Lambda = 0.65$, $\Omega_b = 0.05$, $H_0 = 65 \text{ km s}^{-1} \text{ Mpc}$, $n = 1$ and fluctuation amplitude $\sigma_8 = 0.8$.

galaxies. We assume that the lens galaxies have a constant comoving density (as the lensing rate for an evolving (Press–Schechter) population of lenses differs only by $\lesssim 10\%$ from this case¹⁶) and are primarily early-type (E/S0) SIS galaxies¹⁷ whose population is described by a Schechter function with parameters¹⁸ $n_\star = 0.27 \times 10^{-2} \text{ Mpc}^{-3}$ and $\alpha_s = -0.5$. We assume the Faber–Jackson relation $(L_g/L_{g\star}) = (\sigma_g/\sigma_\star)^4$ where σ_g is the velocity dispersion of the lens galaxy, with $\sigma_\star = 220 \text{ km s}^{-1}$ and a dark matter velocity dispersion that equals the stellar velocity dispersion¹⁷. We ignore dust extinction by the lens galaxy, which should lower¹⁷ the lensing rate by $\lesssim 10\%$. Potential lens galaxies must not be detectable in the survey data used to select the objects. Galaxies having $i^\star < 22.2$ (around 30% of the potential lens population) are not considered part of the lens population. To compute $i^\star$ for a galaxy having velocity dispersion σ at redshift z , we use $L_{g\star}$ from the 2dF early-type galaxy luminosity function¹⁸, the Faber–Jackson relation, colour transformations with a k -correction^{19,20}, and the evolution of the rest-frame B -band mass-to-light ratio²¹.

Our lens model includes microlensing by the population of galactic stars, which is modelled as a de Vaucouleurs profile of point masses embedded in the overall SIS mass distribution. The surface mass density in stars for galaxies at $z = 0$ is normalized so that the total cosmological density parameter of stars²² equals 0.005. At $z > 0$ the total mass density in stars is assumed to be proportional to the cumulative star-formation history^{23,24}. The parameters of the de Vaucouleurs profiles are taken from a study of the fundamental plane²⁵, the microlens mass is $0.1 M_\odot$ and the source size 10^{15} cm (corresponding to ten Schwarzschild radii of a $3 \times 10^8 M_\odot$ black hole). The probabilities $\frac{dP_{\text{sing}}}{d\mu}$ and $\frac{dP_{\text{mult}}}{d\mu_{\text{tot}}}$ closely

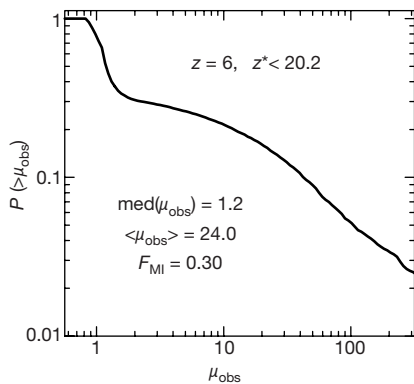


Figure 2 The probability of observing a magnification larger than μ_{obs} for a quasar at a redshift $z = 6$ in a sample with a magnitude limit $z^\star < 20.2$. The distribution is highly skewed, having a median of $\text{med}(\mu_{\text{obs}}) = 1.2$ and a mean of $\langle \mu_{\text{obs}} \rangle = 24.0$. The multiple image fraction is $F_{\text{MI}} = 0.30$. We have also computed the *a posteriori* values of F_{MI} and $\langle \mu \rangle$ for specific quasars. For SDSS 0836-0054 ($z = 5.82$), SDSS 1306-0356 ($z = 5.99$) and SDSS 1030-0524 ($z = 6.28$) we find $F_{\text{MI}} = 0.40, 0.32$ and 0.31 , and $\langle \mu \rangle = 50, 25$ and 23 , respectively. Although Fan *et al.*¹ find that $\beta_h = 3.43$ is consistent with the luminosities of the $z \gtrsim 5.8$ quasars, it is possible that the luminosity function at high redshift is not as steep as $\beta_h = 3.43$. If so, the magnification bias will not be as large and the mean magnification and the fraction of quasars that are multiply imaged will be lower. We have recomputed the lens statistics assuming that the bright end slope is significantly flatter at high redshift. Assuming that $\beta_l = 1.64$ at all redshifts and $\beta_h = 3.43$ for $z < 3$ but $\beta_h = 2.58$ for $z > 3$ (the value found¹¹ for quasars at $z \approx 4.3$) we inferred similar parameters to describe the observed luminosity function as before ($L_\star(0) = 1.5 \times 10^{11} L_\odot$, $z_\star = 1.6$, $\xi = 3.3$, $\zeta = 2.65$). Remarkably, the multiple image fraction is still nearly 0.1 in this case. Additional uncertainty in the calculation arises from the choice of lens model. The value τ_{mult} is proportional to $n_\star \sigma_\star^4$. The dependence of F_{MI} on τ_{mult} is complex (see equation (3)); however, large magnification biases result in a relation that is less sensitive than linear. For example, reducing the value of τ_{mult} by a factor of 1.5 results in $F_{\text{MI}} = 0.22$ if $\beta_h = 3.43$ and $F_{\text{MI}} = 0.043$ if $\beta_h = 2.58$.

resemble the standard form for the SIS and were computed²⁶ by combining numerical magnification maps with the distribution of microlensing optical depth and shear along lensed lines of sight, and the distribution $[\tau_{\text{mult}} \{dP_{\text{mult}}/d\mu\} + (1 - \tau_{\text{mult}}) \{dP_{\text{sing}}/d\mu\}]$ was normalized to have a unit mean.

The fraction of sources which are multiply imaged due to gravitational lensing is

$$F_{\text{MI}}(z) = \frac{\int_0^\infty d\mu' \tau_{\text{mult}} \frac{dP_{\text{mult}}}{d\mu'} N(> \frac{L_{\text{lim}}}{\mu'}, z)}{\int_0^\infty d\mu' \left[\tau_{\text{mult}} \frac{dP_{\text{mult}}}{d\mu'} + (1 - \tau_{\text{mult}}) \frac{dP_{\text{sing}}}{d\mu'} \right] N(> \frac{L_{\text{lim}}}{\mu'}, z)} \quad (2)$$

where $\tau_{\text{mult}} = 0.0059$. We find a value of $F_{\text{MI}} \approx 0.30$ for the colour-selected, flux-limited $z \gtrsim 5.8$ sample. This value is higher by two orders of magnitude than the lens fraction at low redshifts and demonstrates that lensing must already be considered in a sample with only four objects. For comparison, our model predicts $F_{\text{MI}} \approx 0.01$ at $z \approx 2$ for $m_B < 20$. These calculations do not include selection effects for flux ratios or image separations which may serve to lower F_{MI} , but are specific to follow up observations.

A subarcsecond-resolution K-band image has been obtained for the $z = 5.80$ quasar¹, and it was found to be an unresolved point source. To our knowledge, this is the only $z \gtrsim 5.8$ quasar for which subarcsecond-resolution optical imaging is currently available. However, a program to image these quasars with the Hubble Space Telescope that will determine the multiple-image fraction is expected to begin within the coming year (X. Fan, personal communication). We note, however, that single-image quasars might still be magnified by a factor of about 2. Recent Chandra observations²⁷ of the $z = 6.28$ quasar show photons detected on the edge of the extraction ($1.2''$) circle, offering a hint that this quasar may be lensed.

We have computed the distribution of magnifications observed for a sample of quasars brighter than L_{lim} at redshift z :

$$\frac{dP}{d\mu_{\text{obs}}} = \frac{\left[\tau_{\text{mult}} \frac{dP_{\text{mult}}}{d\mu} \Big|_{\mu=\mu_{\text{obs}}} + (1 - \tau_{\text{mult}}) \frac{dP_{\text{sing}}}{d\mu} \Big|_{\mu=\mu_{\text{obs}}} \right] N(> \frac{L_{\text{lim}}}{\mu_{\text{obs}})}}{\int_0^\infty d\mu' \left[\tau_{\text{mult}} \frac{dP_{\text{mult}}}{d\mu} \Big|_{\mu=\mu'} + (1 - \tau_{\text{mult}}) \frac{dP_{\text{sing}}}{d\mu} \Big|_{\mu=\mu'} \right] N(> \frac{L_{\text{lim}}}{\mu'})} \quad (3)$$

In Fig. 2 we show the probability that the magnification of a quasar is higher than μ_{obs} , assuming that the quasar belongs to a sample at a redshift $z \approx 6$ with the SDSS magnitude limit of $z^\star < 20.2$. The distribution is highly skewed; the median magnification is $\text{med}(\mu_{\text{obs}}) \approx 1.2$, whereas the mean is as high as $\langle \mu \rangle = 24$. Thus, one or more of the $z \gtrsim 5.8$ quasars are likely to be highly magnified, whereas most should be magnified at a low level.

The large values of the magnifications and the highly skewed shape of the distributions in Fig. 2 suggest that lensing must alter the observed luminosity function. Indeed, we find that the space density of quasars with $M_B < -27.6$ is increased by 40%, and that the slope is decreased by 0.15. A quasar magnified by μ is detectable to luminosities as low as L_{lim}/μ and has its luminosity, L , overestimated by μ . The factor R_{LD} by which the luminosity density of quasars brighter than L_{lim} is overestimated because of lensing is therefore

$$R_{\text{LD}} = \frac{\int_0^\infty d\mu' \left[\tau_{\text{mult}} \frac{dP_{\text{mult}}}{d\mu'} + (1 - \tau_{\text{mult}}) \frac{dP_{\text{sing}}}{d\mu'} \right] \int_{L_{\text{lim}}/\mu'}^\infty dL' \mu' L' \phi(L')}{\int_{L_{\text{lim}}}^\infty dL' L' \phi(L')} \quad (4)$$

We find $R_{\text{LD}} \approx 2$, implying that naive computation of the quasar luminosity density from the $z \gtrsim 5.8$ sample might significantly overestimate its true value. Magnification bias also affects quasars that are not multiply imaged. We therefore predict an enhanced angular correlation on the sky between $z \approx 6$ quasars and foreground galaxies.

Traces of neutral hydrogen at high redshift absorb flux just blueward of the Ly α resonance. The resulting Gunn–Peterson²⁸ trough has been observed in the spectrum of the highest redshift quasar^{7,29}, and is the primary indicator of the reionization epoch. We find that about 40% of multiple image lens galaxies ($i^* < 22.2$) contribute flux in the Gunn–Peterson trough above the current upper limit. Contamination of the Gunn–Peterson trough in lensed quasar spectra will therefore limit their use as a probe of the epoch of reionization.

The high source redshifts imply image separations that are slightly larger than usual¹⁶, about 1–2". In addition, the lenses are likely to be found at higher redshift¹⁶ owing in part to the higher source redshifts but also because bright (low- z) lenses are excluded. Lensing of high-redshift quasars allows measurement of the masses of the lens galaxies²¹ at redshifts higher than is currently possible. Furthermore, quasar microlensing should be common over a ten-year base line²⁶. This offers the exciting possibility of measuring the source size, and hence black hole mass, which in turn yields the ratio between quasar luminosity and its Eddington value, a quantity that will be useful in constraining models of structure formation³. □

Received 20 February; accepted 26 April 2002; doi:10.1038/nature00794.

1. Fan, X. *et al.* Survey of $z > 5.8$ quasars in the Sloan digital sky survey. I. Discovery of three new quasars and the spatial density of luminous quasars at $z \sim 6$. *Astron. J.* **122**, 2833–2849 (2001).
2. Turner, E. L. Quasars and galaxy formation. I—The Z greater than 4 objects. *Astron. J.* **101**, 5–17 (1991).
3. Haiman, Z. & Loeb, A. What is the highest plausible redshift for quasars? *Astrophys. J.* **503**, 505–517 (2001).
4. Djorgovski, S. G., Castro, S., Stern, D. & Mahabal, A. A. On the threshold of the reionization epoch. *Astrophys. J.* **560**, L5–L8 (2001).
5. Press, W. H. & Schechter, P. Formation of galaxies and clusters of galaxies by self-similar gravitational condensation. *Astrophys. J.* **187**, 425–438 (1974).
6. Barkana, R. & Loeb, A. In the beginning: the first sources of light and the reionization of the universe. *Phys. Rep.* **349**, 125–238 (2001).
7. Becker, R. H. *et al.* Evidence for reionization at $z \sim 6$: Detection of a Gunn–Peterson trough in a $z = 6.28$ quasar. *Astron. J.* **122**, 2850–2857 (2001).
8. Madau, P., Haardt, F. & Rees, M. J. Radiative transfer in a clumpy universe. III. The nature of cosmological ionizing sources. *Astrophys. J.* **514**, 648–659 (1999).
9. Haiman, Z. & Loeb, A. Observational signatures of the first quasars. *Astrophys. J.* **503**, 505–517 (1998).
10. Pei, Y. C. The luminosity function of quasars. *Astrophys. J.* **438**, 623–631 (1995).
11. Fan, X. *et al.* High redshift quasars found in the Sloan digital sky survey commissioning data. IV. Luminosity function from the full equatorial sample. *Astron. J.* **121**, 54–65 (2001).
12. Møller, P. & Jakobsen, P. The Lyman continuum opacity at high redshifts: through the Lyman forest and beyond the Lyman valley. *Astron. Astrophys.* **228**, 299–309 (1990).
13. Ostriker, J. P. & Vietri, M. The statistics of gravitational lensing. III Astrophysical consequences of quasar lensing. *Astrophys. J.* **300**, 68–76 (1986).
14. Turner, E. L., Ostriker, J. P. & Gott, R. The statistics of gravitational lenses: The distributions of image angular separations and lens redshifts. *Astrophys. J.* **284**, 1–22 (1984).
15. Turner, E. L. Gravitational lensing limits on the cosmological constant in a flat universe. *Astrophys. J.* **365**, L43–L46 (1990).
16. Barkana, R. & Loeb, A. High-redshift quasars: Their predicted size and surface brightness distributions and their gravitational lensing probability. *Astrophys. J.* **531**, 613–623 (2000).
17. Kochanek, C. S. Is there a cosmological constant? *Astrophys. J.* **466**, 638–659 (1996).
18. Madgwick, D. S. *et al.* The 2dF galaxy redshift survey: Galaxy luminosity functions per spectral type. Preprint astro-ph/0107197 at (<http://xxx.lanl.gov>) (2001).
19. Blanton, M. R. *et al.* The luminosity function of galaxies in SDSS commissioning data. *Astron. J.* **121**, 2358–2380 (2001).
20. Fukugita, M., Shimasaku, K. & Ichikawa, T. Galaxy colours in various photometric band systems. *Pub. R. Astron. Soc. Pacif.* **107**, 945–958 (1995).
21. Koopmans, L. V. E. & Treu, T. The stellar velocity dispersion of the lens galaxy in MG2016 + 112 at $z = 1.004$. *Astrophys. J.* **568**, L5–L8 (2002).
22. Fukugita, M., Hogan, C. J. & Peebles, P. J. E. The cosmic baryon budget. *Astrophys. J.* **503**, 518–530 (1998).
23. Hogg, D. A meta-analysis of cosmic star-formation history. Preprint astro-ph/105280 at (<http://xxx.lanl.gov>) (2001).
24. Nagamine, K., Cen, R. & Ostriker, J. P. Luminosity density of galaxies and cosmic star-formation rate from Λ cold dark matter hydrodynamical simulations. *Astrophys. J.* **541**, 25–36 (2000).
25. Djorgovski, S. & Davis, M. Fundamental properties of elliptical galaxies. *Astrophys. J.* **313**, 59–68 (1987).
26. Wyithe, J. S. B. & Turner, E. L. Cosmological microlensing statistics: Variability rates for quasars and GRB afterglows, and implications for macrolensing magnification bias and flux ratios. Preprint astro-ph/0203214 at (<http://xxx.lanl.gov>) (2002).
27. Schwartz, D. A. A Chandra search for jets in redshift 6 quasars. Preprint astro-ph/0202190 at (<http://xxx.lanl.gov>) (2002).
28. Gunn, J. E. & Peterson, B. A. On the density of neutral hydrogen in intergalactic space. *Astrophys. J.* **142**, 1633–1641 (1965).
29. Pentericci, L. *et al.* VLT optical and near-IR observations of the $z = 6.28$ quasar SDSS 1030 + 0524. Preprint astro-ph/0112075 at (<http://xxx.lanl.gov>) (2001).
30. Ferrarese, L. Beyond the bulge: a fundamental relation between supermassive black holes and dark matter halos. Pre-print astro-ph/0203469 at (<http://xxx.lanl.gov>) (2001).

Acknowledgements

We thank E. Turner, J. Winn and R. Barkana for discussions. This work was supported in part by grants from the NSF and NASA. J.S.B.W. is supported by a Hubble Fellowship.

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to A.L. (e-mail: aloeb@cfa.harvard.edu).

The nature and transport mechanism of hydrated hydroxide ions in aqueous solution

Mark E. Tuckerman*, Dominik Marx† & Michele Parrinello‡§

* Department of Chemistry and Courant Institute of Mathematical Sciences, New York University, 4 Washington Place, New York, New York 10003, USA

† Lehrstuhl für Theoretische Chemie, Ruhr-Universität Bochum, 44780 Bochum, Germany

‡ Max-Planck-Institut für Festkörperforschung, Heisenbergstrasse 1, 70569 Stuttgart, Germany

Compared to other ions, protons (H^+) and hydroxide ions (OH^-) exhibit anomalously high mobilities in aqueous solutions¹. On a qualitative level, this behaviour has long been explained by ‘structural diffusion’—the continuous interconversion between hydration complexes driven by fluctuations in the solvation shell of the hydrated ions. Detailed investigations have led to a clear understanding of the proton transport mechanism at the molecular level^{2–8}. In contrast, hydroxide ion mobility in basic solutions has received far less attention^{2,3,9,10}, even though bases and base catalysis play important roles in many organic and biochemical reactions and in the chemical industry. The reason for this may be attributed to the century-old notion¹¹ that a hydrated OH^- can be regarded as a water molecule missing a proton, and that the transport mechanism of such a ‘proton hole’ can be inferred from that of an excess proton by simply reversing hydrogen bond polarities^{11–18}. However, recent studies^{2,3} have identified OH^- hydration complexes that bear little structural similarity to proton hydration complexes. Here we report the solution structures and transport mechanisms of hydrated hydroxide, which we obtained from first-principles computer simulations that explicitly treat quantum and thermal fluctuations of all nuclei^{19–21}. We find that the transport mechanism, which differs significantly from the proton hole picture, involves an interplay between the previously identified hydration complexes^{2,3} and is strongly influenced by nuclear quantum effects.

Transport of a hydrated proton, H_3O^+ , in water involves a continuous interconversion between two hydration complexes (Fig. 1a–c). A tricoordinate $H_3O^+(H_2O)_3$ complex (Fig. 1a) is transformed via proton transfer into a $[H_2O \cdots H \cdots OH_2]^+$ or $H_5O_2^+$ complex (Fig. 1b). This process, known as the ‘structural diffusion’ or ‘Grotthuss’ mechanism¹, is driven by fluctuations in the second solvation shell of H_3O^+ , which reduce the coordination of a first-solvation-shell water from four to three^{2–5}. This water molecule is, thus, ‘prepared’ to become a properly solvated H_3O^+ by proton transfer $H_3O_a^+ \cdots H_2O_b \rightleftharpoons H_2O_a \cdots H_3O_b^+$, where O_a and O_b are the two oxygens involved in the hydrogen bond, respectively.

§ Present address: CSCS-Swiss Center for Scientific Computing, Via Cantonale, 6928 Manno, Switzerland; Department of Physical Chemistry, ETH, Höggerberg HCI, 8093 Zürich, Switzerland.

Owing to quantum effects, the actual defect structure is best described as a fluxional defect⁵. Applying the proton-hole concept to the OH^- case, one would expect the occurrence of tricoordinate $\text{OH}^-(\text{H}_2\text{O})_3$ (in which the hydroxide oxygen accepts three hydrogen bonds) and H_3O_2^- (that is $[\text{HO}\cdots\text{H}\cdots\text{OH}]^-$) complexes, analogous to $\text{H}_3\text{O}^+(\text{H}_2\text{O})_3$ and H_5O_2^+ , respectively. Moreover, structural diffusion would be driven by similar second-solvation-shell fluctuations. Interpretations of Raman and infrared spectra of concentrated basic solutions have been guided by the proton-hole picture, a fact that has led to controversies about the structure and

dynamics of hydrated hydroxide^{14–17}. Prompted by this controversy, we undertook an extensive *ab initio* path integral^{19,20} study of a hydroxide ion in bulk water at room temperature (see Fig. 2 for details).

For each configuration generated in the simulation, the OH^- charge defect, designated $(\text{O}^*\text{H}')^-$, is located in the hydrogen-bond (H-bond) network by identifying the oxygen with a single hydrogen. Next, for each H-bond, a displacement coordinate, $\tilde{\delta} = R_{\text{O}_a\text{H}} - R_{\text{O}_b\text{H}}$, is defined, where $R_{\text{O}_a\text{H}}$ and $R_{\text{O}_b\text{H}}$ are the distances between the shared proton and the two oxygens. A small value of $\tilde{\delta}$ indicates a propensity for proton transfer. Finally, a transfer coordinate, δ , is defined in each configuration by selecting the H-bond with the smallest $\tilde{\delta}$. This H-bond, designated $\text{O}^*\cdots\text{H}^*\bar{\text{O}}$, is considered to be the ‘most active’ or most likely to experience proton transfer⁵. Invariably, it is found that one of the oxygens in this H-bond corresponds to the defect, O^* .

In order to identify the role played by various solvation complexes of OH^- , configurations with large and small $|\delta|$ are selected. Figure 2a shows that, at large $|\delta|$, O^* accepts, on average, four H-bonds from neighbouring waters. Representative configurations

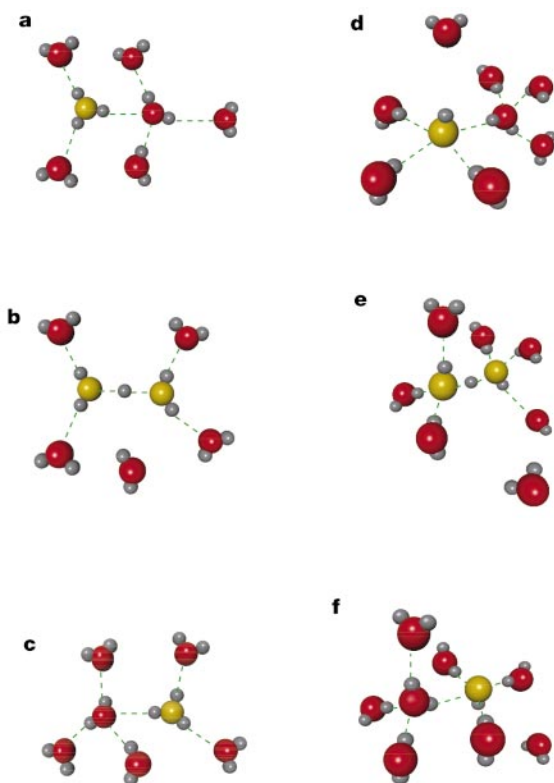


Figure 1 Schematic illustration of proton and hydroxide ion transport in water at room temperature. Panels **a–c** show the transport mechanisms of H_3O^+ , and panels **d–f** the transport mechanisms of OH^- . Red and grey spheres are oxygen and hydrogen atoms, respectively, and yellow spheres are the H_3O^+ and OH^- defect oxygens. Dashed green lines denote hydrogen bonds. These simplified sketches are idealized renderings of the classical limit of the ‘real’ processes shown in Fig. 3 and Fig. 1 of ref. 5. H_3O^+ mechanism: In **a**, the H_3O^+ is shown in a threefold-coordinated, $\text{H}_3\text{O}^+(\text{H}_2\text{O})_3$, state, and a complete coordination shell is shown for one of the first-solvation-shell waters. Thermal fluctuations cause an H-bond breaking between a first and second solvation-shell water of H_3O^+ to occur (**b**), leaving the first-solvation-shell water with a threefold coordination pattern, followed by a shrinking of the first-shell O^*O distance and a migration of the proton to the centre of the bond, forming an intermediate H_5O_2^+ complex. Complete transfer of the proton (**c**), yields a properly solvated H_3O^+ at a new site in the H-bond network. The ‘gate closes’ (no further proton transfer events are possible until another H-bond breaking event occurs) if the newly formed water acquires a fourth H-bond (shown as the water released by the H-bond breaking event in **b**, although it could be any nearby solvent water). OH^- mechanism: In **d**, the OH^- is shown in a fourfold-coordinated, $\text{OH}^-(\text{H}_2\text{O})_4$, state, and a complete coordinate shell is shown for one of the first-solvation-shell waters. A neighbouring solvent is shown above the OH^- hydrogen. Thermal fluctuations cause a first-solvation-shell H-bond breaking event (**e**), leaving the OH^- in a threefold $\text{OH}^-(\text{H}_2\text{O})_3$ state, followed by a shrinking of a first-shell O^*O distance and the formation of a weak H-bond between the OH^- hydrogen and a neighbouring solvent water (see, also, Fig. 3b–d). Complete transfer of the proton (**f**) causes the $\text{OH}^-(\text{H}_2\text{O})_3$ to migrate to a new site in the H-bond network. The ‘gate closes’ if the newly formed hydroxide acquires a fourth H-bond (shown as the water released in **b**, although it could be any nearby solvent water), to complete the $\text{OH}^-(\text{H}_2\text{O})_4$ structure.

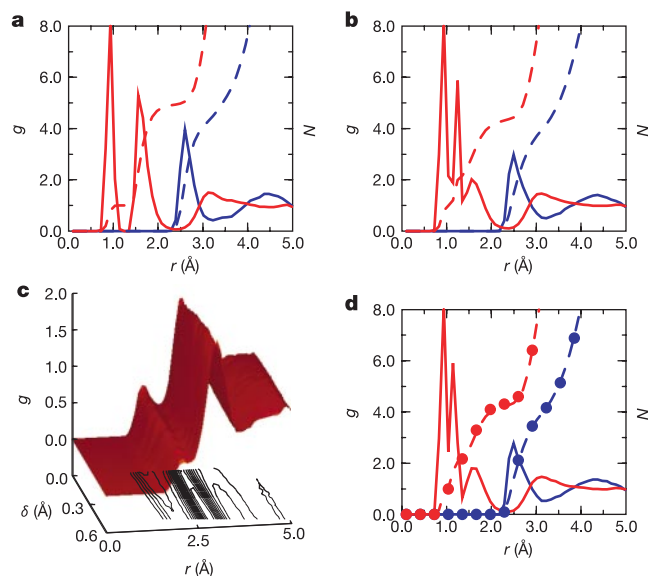


Figure 2 Radial distribution functions and coordination numbers during proton transfer. Shown are radial distribution functions, $g(r)$, and, on the same scale, running coordination numbers, $N(r)$, with respect to the hydroxide ion and a first-solvation-shell water at different stages in the proton transfer process. **a**, Radial distribution functions, $g_{\text{O}^*\text{O}}$ (solid blue) and $g_{\text{O}^*\text{H}}$ (solid red) and coordination numbers, $N_{\text{O}^*\text{O}}$ (dashed blue) and $N_{\text{O}^*\text{H}}$ (dashed red), with respect to the OH^- , O^* , for values of the proton transfer coordinate $|\delta| = R_{\text{O}^*\text{H}^*} - R_{\text{O}^*\text{H}} > 0.5 \text{ \AA}$. Here, H^* denotes the shared proton, and $\bar{\text{O}}$ denotes the oxygen of a water molecule that donates the ‘most active’ H-bond to O^* , that is, $\text{O}^*\cdots\text{H}^*\bar{\text{O}}$. **b**, As **a** except for $|\delta| < 0.1 \text{ \AA}$. **c**, Two-dimensional generalization of a hydrogen-oxygen radial distribution function $g_{\text{H}^*\text{O}}(r, |\delta|)$ with respect to the OH^- hydrogen, H^* , as a function of r and $|\delta|$. **d**, Radial distribution functions and coordination numbers with respect to $\bar{\text{O}}$ for $|\delta| < 0.1 \text{ \AA}$. The colour and line types are the same as in **a**. Coordination numbers from **b** are superimposed with red (O^*H) and blue (O^*O) circles. The simulations were performed using CPMD Version 3.0 (J. Hutter *et al.*, Max-Planck-Institut für Festkörperforschung und IBM Zurich Research Laboratory 1995–1996)²¹ at ambient conditions using 31 H_2O with one OH^- anion in a periodic cubic box of side length $9.8692 \text{ \AA}$, employing the BLYP functional, Troullier-Martins type pseudopotentials, and a Γ -point plane wave expansion of the valence orbitals up to 70 Ry . This particular scheme has been shown to yield a good description of dynamical processes in aqueous systems^{5,6,30,32}. The path integral was discretized using $P = 8$ Trotter replicas following the methodology of ref. 20. After careful equilibration, quantum and classical nuclear trajectories consisting of 322,000 and 346,000 configurations, with time steps of 5.0 a.u. and 7.0 a.u. , respectively, were generated via the Car-Parrinello algorithm²².

(Fig. 3a) reveal that the accepted H-bonds are in a roughly planar arrangement, forming the $\text{OH}^-(\text{H}_2\text{O})_4$ complex. Previous solution phase² and quantum-chemical²³ calculations have confirmed the stability of this structure. Furthermore, a recent spectroscopic study indicates that there is no cationic analogue of this complex²⁴.

Importantly, it is also found that in roughly half of the configurations, the hydroxide hydrogen, H' , donates an additional weak H-bond, giving a total O^*O coordination number of $N_{\text{O}^*\text{O}} \approx 4.5$.

For small $|\delta|$, the O^*O peak in the radial distribution functions (Fig. 2b) shifts to smaller distances, and $N_{\text{O}^*\text{O}}$ decreases from 4.5 to

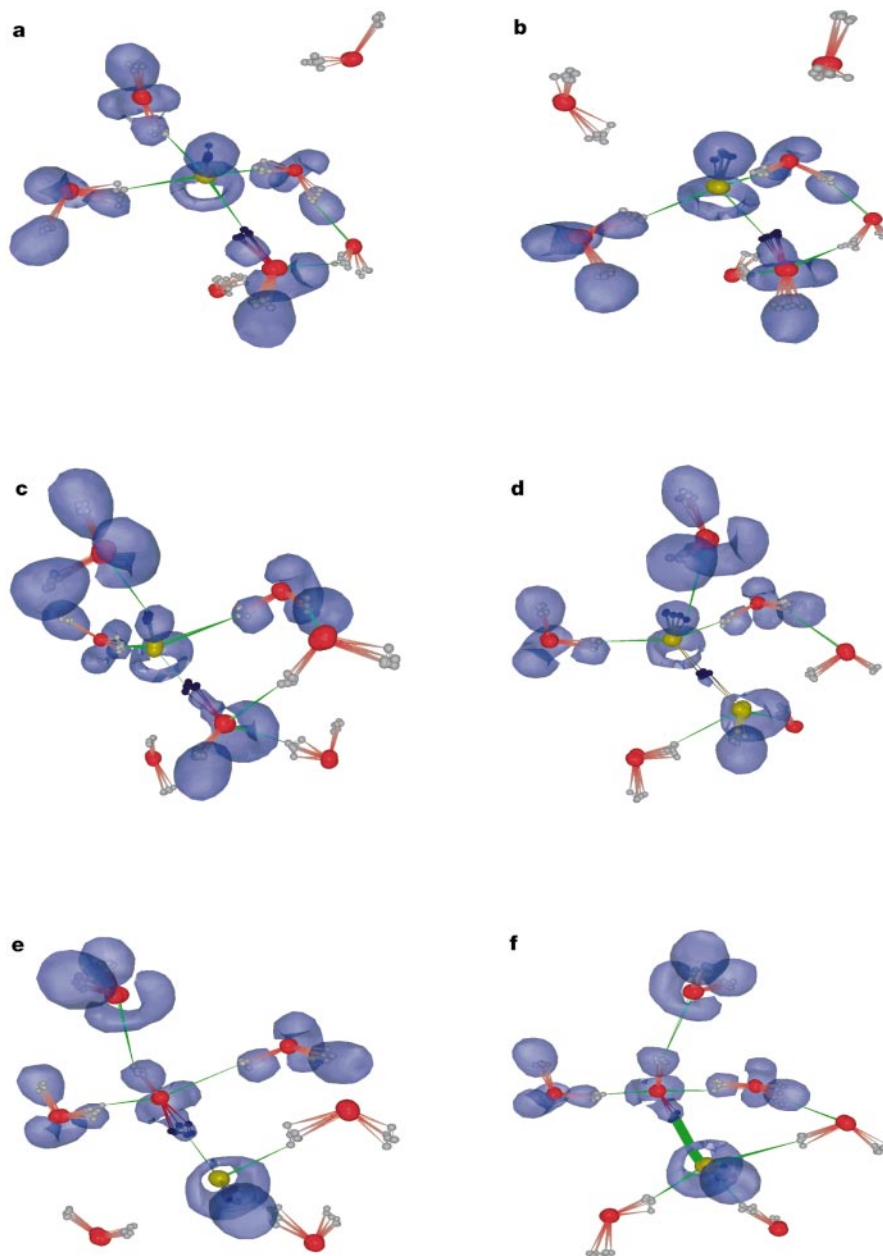


Figure 3 Representative configurations showing the proton transfer mechanism. Each panel shows quantum-mechanical particle density snapshots of the OH^- together with relevant first and second solvation-shell water molecules. Oxygen and hydrogen atoms are shown as red and grey spheres, respectively. The defect oxygen, O^* and shared proton in the most active H-bond are shaded yellow and black, respectively. H-bonds are schematically shown as green lines. In blue are shown the $\eta = 0.93$ isosurfaces of the electron localization function (ELF)²⁹ of the path centroid configuration for the OH^- and only those water molecules which are H-bonded to the OH^- . **a**, The OH^- is in its inactive $\text{OH}^-(\text{H}_2\text{O})_4$ structure with fourfold planar coordination around O^* . The ELF shows that the three lone pairs around O^* are in a delocalized ring structure, thus supporting the hypercoordination of the OH^- (see text). An additional water molecule, which participates in the activation process (**c** and **d**), appears in the upper right corner. **b**, A first-solvation-shell H-bond breaking event occurs, which transforms the $\text{OH}^-(\text{H}_2\text{O})_4$ structure into an approximately tetrahedral $\text{OH}^-(\text{H}_2\text{O})_3$ structure. **c**, A weak H-bond between the OH^-

hydrogen, H' , and the bulk water molecule from **a** and **b** is formed. As shown, when this H-bond forms, one of the shared protons between O^* and a coordinating water begins to transfer. **d**, As the proton is transferred, the environments around each OH moiety become similar (see, also, Fig. 2d). (Note that a complete solvation shell for the water that transfers its proton to O^* is not shown. However, the coordination number around this water is, nevertheless, 4.0—see Fig. 2d). The ELF shows two distorted rings around the yellow oxygens, emphasizing the symmetry of this configuration. **e**, After the proton transfer, the original OH^- has been transformed into a properly solvated water molecule with two lone pairs, whereas the newly formed OH^- (now shown in yellow) possesses the ring-shaped ELF. **f**, The process is completed by the acceptance of a fourth H-bond from the newly formed O^* , which ‘closes the gate’ by forming a new, inactive $\text{OH}^-(\text{H}_2\text{O})_4$ complex with a fully intact ELF ring as in **a**. Although *ab initio* path-integral molecular dynamics method does not rigorously produce real-time evolution, processes that must occur and their approximate sequence can be inferred qualitatively.

4.1, very close to the value, 4.0, obtained for bulk water molecules. The number of accepted and donated H-bonds is 3.2 and 0.9, respectively. These imply that the emergent structure during proton transfer is the tetrahedral $\text{OH}^-(\text{H}_2\text{O})_3$ complex in which O^* is approximately threefold coordinated (Fig. 3b), and that the donated H-bond involving H' plays a crucial role in activating the proton transfer process (Fig. 3c, d). The importance of this $\text{H}'\cdots\text{O}$ H-bond is further demonstrated in the two-dimensional radial distribution function, $g_{\text{H}'\text{O}}(r, |\delta|)$ of Fig. 2c. A weak signal at $r \approx 2 \text{ \AA}$, first apparent at large $|\delta|$, becomes prominent as $|\delta| \rightarrow 0$. It should be noted that the location of the O^*O peak of the δ -averaged radial distribution function, $2.5 \text{ \AA}$, is in good agreement with recent neutron diffraction data²⁵. Moreover the δ -averaged O^*O coordination, 4.3, is consistent with experimental estimates²⁶ (although the "effective hydration number" in ref. 26 is not precisely equivalent to the microscopic coordination number).

The above findings have led us to propose an alternative mechanism to that of the proton-hole picture (Fig. 3). Activation of a proton transfer event requires that the solvation pattern around OH^- closely resemble that of a bulk water molecule. This pattern (which is not a free-energy minimum), is realized by two concerted yet distinct processes. The first is a transformation from the inactive $\text{OH}^-(\text{H}_2\text{O})_4$ to the active $\text{OH}^-(\text{H}_2\text{O})_3$ solvation state, a process that involves the breaking of a H-bond in the first solvation shell of OH^- (as shown in the change from Fig. 3a to Fig. 3b). The relatively facile conversion from $\text{OH}^-(\text{H}_2\text{O})_4$ to $\text{OH}^-(\text{H}_2\text{O})_3$ is consistent with the estimated activation energy, $1.16 \text{ kcal mol}^{-1}$, reported in ref. 23 and with recent spectroscopic data²⁴. The importance of this step can be understood from the chemical differences between the $\text{OH}^-(\text{H}_2\text{O})_4$ and $\text{OH}^-(\text{H}_2\text{O})_3$ complexes. In the former, proton transfer through a H-bond would transform the OH^- defect into a H_2O molecule that accepts three H-bonds, an unfavourable configuration. In the latter, proton transfer through one of the acceptor H-bonds produces a H_2O with the correct coordination. The second process required to initiate proton migration is the formation of a weak H-bond between the hydroxide hydrogen, H' , and a neighbouring water molecule (Fig. 3b to c). When this occurs in the $\text{OH}^-(\text{H}_2\text{O})_3$ complex, the OH^- is (fourfold) coordinated in such a way that when the proton, H' , is transferred, the original $(\text{O}^*\text{H}')^-$ becomes a tetrahedrally solvated water molecule with an ideal bulk solvation pattern. Note that, in the proton-hole picture, this $\text{H}'\cdots\text{O}$ H-bond (Figs 2c, 3c, 3d) is assumed never to form.

Once the $\text{OH}^-(\text{H}_2\text{O})_3$ complex is activated, the environment around the two oxygens in the most active H-bond is nearly

identical (Fig. 2d). This observation suggests that, for $|\delta| \approx 0$, the structural motif of a solvated H_3O_2^- core, emerges. However, in contrast to the proton-hole scenario, here, the H_3O_2^- is a transient state that is necessarily visited when the proton is transferred. Thus, the complete structural diffusion mechanism, depicted in Fig. 3, can be summarized as: (1) transformation of $\text{OH}^-(\text{H}_2\text{O})_4$ to $\text{OH}^-(\text{H}_2\text{O})_3$ by first-solvation-shell H-bond breaking (Fig. 3a to b); (2) activation of $\text{OH}^-(\text{H}_2\text{O})_3$ by formation of the $\text{H}'\cdots\text{O}$ H-bond (Fig. 3b to c); (3) formation of the transient H_3O_2^- complex by partial proton transfer (Fig. 3c to d); (4) completion of proton transfer resulting in a new O^* and $\text{OH}^-(\text{H}_2\text{O})_3$ complex (Fig. 3d to e); (5) acceptance of a fourth H-bond by the new O^* resulting in a change from $\text{OH}^-(\text{H}_2\text{O})_3$ to $\text{OH}^-(\text{H}_2\text{O})_4$ and consequent inactivation (Fig. 3e to f). This completes the cycle: the $\text{OH}^-(\text{H}_2\text{O})_4$ complexes in Figs 3f and 3a are isostructural but are located at different sites in the network. In Fig. 1d–f a simplified sketch of this mechanism is contrasted to that for structural diffusion of H_3O^+ in acids (Fig. 1a–c).

Having established the mechanism, we examine how closely a purely classical treatment of the nuclear motion at 300 K approximates this picture. Classically (Fig. 4a), there is a significant funnelling of the angle (θ) between the $\text{O}^*\text{H}'$ and O^*O bonds about the HOH bulk water bend angle ($\sim 104^\circ$) for $|\delta| \approx 0$, implying that the $\text{OH}^-(\text{H}_2\text{O})_3$ complex must be close to its ideal tetrahedral geometry. However, when nuclear quantum effects are included (Fig. 4b), this funnelling is largely washed out, implying a significant probability of proton transfer even when the geometry of the complex is distorted. Such a pronounced 'corner cutting' quantum effect²⁷, having no analogue in hydrated H_3O^+ (ref. 5), may also explain why the observed H/D isotope effect on the mobility is larger in basic than in acidic solutions¹⁸.

Nuclear quantum effects also lower the proton transfer free energy barrier along the δ -coordinate from roughly 1.28 to $0.34 \text{ kcal mol}^{-1}$ (Fig. 4c). Despite this reduction, the free energy retains its double-well character, implying that H_3O_2^- is a transient complex or activated state (Figs 2d and 3d). This contradicts the proton-hole picture, in which H_3O_2^- is understood to be a stable complex. It is also in contrast to the hydrated H_3O^+ case, in which a smaller ($0.6 \text{ kcal mol}^{-1}$) barrier obtained at the classical level is washed out by nuclear quantum effects⁵. Finally, it contrasts with the gas-phase studies in ref. 28, where a pronounced barrier in the H_3O_2^- complex is also completely washed out by nuclear quantum effects. If the present proton transfer free-energy value, $0.34 \text{ kcal mol}^{-1}$, is combined with the $1.16 \text{ kcal mol}^{-1}$ (ref. 23)

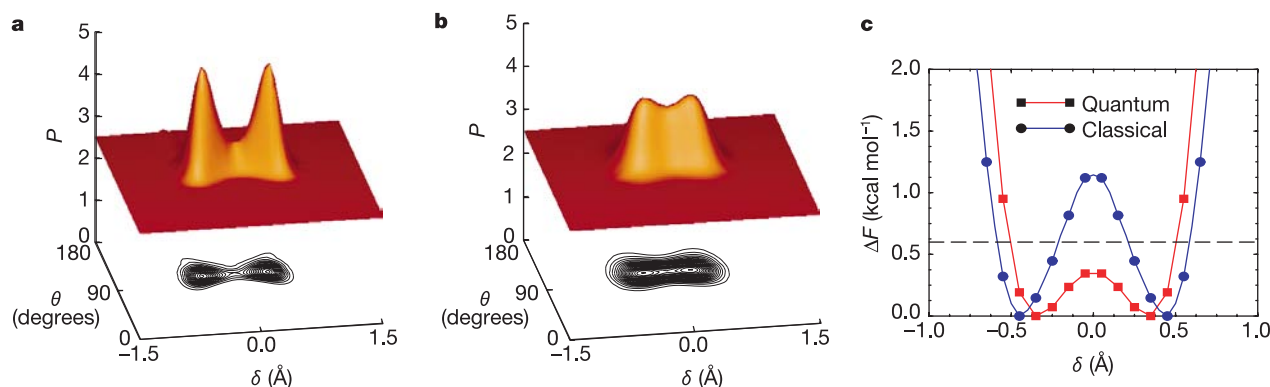


Figure 4 Structural changes occurring during proton transfer. Shown are the normalized two-dimensional probability distribution function, $P(\delta, \theta)$, of the proton displacement coordinate, δ and the angle, θ , between the OH^- covalent bond axis and the O^*O vector of the most active H-bond for the classical (a) and quantum (b) simulations at 300 K. Contours of these distribution functions are also shown. The distribution functions are

smoothed for presentation and symmetrized about $\delta = 0 \text{ \AA}$. Comparison of the free-energy profiles (c) along the δ coordinate for classical (circles) and quantum (squares) treatment of the nuclei. Note the centroid of the coordinate, δ , has been used for the quantum profile.

required for step (1) of the above mechanism, and assuming an additional $0.5 \text{ kcal mol}^{-1}$ (ref. 18) to reduce the length of each of the three H-bonds in the $\text{OH}^-(\text{H}_2\text{O})_3$ complex by $0.1 \text{ \AA}$ after step (1), an estimate of approximately 3 kcal mol^{-1} is obtained for the total activation energy of our mechanism. This rough estimate is consistent with the experimental value obtained from hydroxide mobility data¹⁸.

The unexpectedly high coordination of the solvated hydroxide^{2,23,26} contrasts with the textbook picture that each of the three lone pairs of O^* accepts a single H-bond. An explanation of this phenomenon is obtained by analysing the electron localization function²⁹ along the mechanistic path of Fig. 3. The electron localization function indicates spatial regions where electron pairs are most likely to be found. The most striking feature of these functions (Fig. 3) is the presence of a continuous ring around the OH^- bond axis^{30,31}. This finding is corroborated by its electrostatic potential that displays a ring of maximum negative charge around its O terminus. This pattern, which suggests a lack of directionality in H-bonding to O^* , supports both structural flexibility and high coordination, and is consistent with the spectroscopic picture of ref. 24.

The analysis presented here emphasizes the complex chemical differences between the hydrated H_3O^+ and OH^- ions, and their effect on the anomalous transport mechanisms in acidic and basic solution (Fig. 1). We recognize the temptation to invoke symmetry arguments as a method to intuit mechanisms of seemingly similar processes. However, this approach cannot yield a correct picture when the chemistry of the two species is sufficiently different. Nevertheless, a common pattern of charge migration in acids and bases emerges: both require that the nascent species in the proton transfer process be correctly 'pre-solvated' by way of specific fluctuations in the local H-bond network. □

Received 4 January; accepted 10 May 2002; doi:10.1038/nature00797.

- Atkins, P. W. *Physical Chemistry* 6th edn 740–741 (Oxford Univ. Press, Oxford, 1998).
- Tuckerman, M. E., Laasonen, K., Sprik, M. & Parrinello, M. *Ab initio* molecular dynamics simulation of the solvation and transport of H_3O^+ and OH^- ions in water. *J. Phys. Chem.* **99**, 5749–5752 (1995).
- Tuckerman, M. E., Laasonen, K., Sprik, M. & Parrinello, M. *Ab initio* molecular dynamics simulation of the solvation and transport of hydronium and hydroxide ions in water. *J. Chem. Phys.* **103**, 150–161 (1995).
- Agmon, N. The Grotthuss mechanism. *Chem. Phys. Lett.* **244**, 456–462 (1995).
- Marx, D., Tuckerman, M. E., Hutter, J. & Parrinello, M. The nature of the hydrated excess proton in water. *Nature* **397**, 601–604 (1999).
- Marx, D., Tuckerman, M. E. & Parrinello, M. Solvated excess protons in water: quantum effects on the hydration structure. *J. Phys. Condens. Matter* **12**, A153–A159 (2000).
- Vuilleumier, R. & Borgis, D. Transport and spectroscopy of the hydrated proton: A molecular dynamics study. *J. Chem. Phys.* **111**, 4251–4266 (1999).
- Day, T. J. F., Schmitt, U. W. & Voth, G. A. The mechanism of hydrated proton transport in water. *J. Am. Chem. Soc.* **122**, 12027–12028 (2000).
- Tuñón, I., Rinaldi, D., Ruiz-López, M. F. & Rivail, J. L. Hydroxide ion in liquid water: Structure, energetics, and proton transfer using a mixed discrete-continuum *ab initio* model. *J. Phys. Chem.* **99**, 3798–3805 (1995).
- Muller, R. P. & Warshel, A. *Ab initio* calculations of free energy barriers for chemical reactions in solution. *J. Phys. Chem.* **99**, 17516–17524 (1995).
- Hückel, E. Theorie der Beweglichkeiten des Wasserstoff- und Hydroxylions in wässriger Lösung. *Z. Elektrochem.* **34**, 546–562 (1928).
- Eigen, M. Proton transfer, acid-base catalysis, and enzymatic hydrolysis. Part I: Elementary processes. *Angew. Chem. Int. Edn* **3**, 1–19 (1964).
- Stillinger, F. H. in *Theoretical Chemistry: Advances and Perspectives* 177–234 (eds Eyring, H. & Henderson, D.) 177–234 (Academic, New York, 1978).
- Zatsepina, G. N. State of the hydroxide ion in water and in aqueous solution. *J. Struct. Chem.* **12**, 894–898 (1971).
- Schiöberg, D. & Zundel, G. Very polarizable hydrogen bonds in solution of bases having infra-red absorption continua. *J. Chem. Soc. Faraday Trans. II* **69**, 771–781 (1973).
- Librovich, N. B., Sakun, V. P. & Sokolov, N. D. H^+ and OH^- ions in aqueous solutions—Vibrational spectra of hydrates. *Chem. Phys.* **39**, 351–366 (1979).
- Khoshtariya, D. E. & Berdzenishvili, N. O. A new dynamic elementary act model for thermal and photoinduced proton self-exchange through the lyate ion hydrogen bridges in solutions. *Chem. Phys. Lett.* **196**, 607–613 (1992).
- Agmon, N. Mechanism of hydroxide mobility. *Chem. Phys. Lett.* **319**, 247–252 (2000).
- Marx, D. & Parrinello, M. *Ab initio* path-integral molecular dynamics. *Z. Phys. B* **95**, 143–144 (1994).
- Tuckerman, M. E., Marx, D., Klein, M. L. & Parrinello, M. Efficient and general algorithms for path integral Car-Parrinello molecular dynamics. *J. Chem. Phys.* **104**, 5579–5588 (1996).
- Marx, D. & Hutter, J. in *Modern Methods and Algorithms of Quantum Chemistry* (ed. Grotenhorst, J.) 301–449 (John von Neumann Institute for Computing, Forschungszentrum Jülich, 2000); also at (<http://www.theochem.ruhr-uni-bochum.de/go/cprev.html>) (2000).

- Car, R. & Parrinello, M. Unified approach for molecular dynamics and density-functional theory. *Phys. Rev. Lett.* **55**, 2471–2474 (1985).
- Novoa, J. J., Mota, F., del Valle, C. P. & Planas, M. Structure of the first solvation shell of the hydroxide anion. A model study using $\text{OH}^-(\text{H}_2\text{O})_n$ ($n = 4, 5, 6, 11, 17$) clusters. *J. Phys. Chem. A* **101**, 7842–7853 (1997).
- Chaudhuri, C. *et al.* Infrared spectra and isomeric structures of hydroxide ion-water clusters $\text{OH}^-(\text{H}_2\text{O})_{1-5}$: a comparison with $\text{H}_3\text{O}^+(\text{H}_2\text{O})_{1-5}$. *Mol. Phys.* **99**, 1161–1173 (2001).
- Bruni, F., Ricci, M. A. & Soper, A. K. Structural characterization of NaOH aqueous solution in the glass and liquid states. *J. Chem. Phys.* **114**, 8056–8063 (2001).
- Buchner, R., Hefter, G., May, P. M. & Sipos, P. Dielectric relaxation of dilute aqueous NaOH, NaAl(OH)₄ and NaB(OH)₄. *J. Phys. Chem. B* **103**, 11186–11190 (1999).
- Truhlar, D. G. & Kupperman, A. Exact tunneling calculations. *J. Am. Chem. Soc.* **93**, 1840–1851 (1971).
- Tuckerman, M. E., Marx, D., Klein, M. L. & Parrinello, M. On the quantum nature of the shared proton in hydrogen bonds. *Science* **275**, 817–820 (1997).
- Becke, A. D. & Edgecombe, K. E. A simple measure of electron localization in atomic and molecular systems. *J. Chem. Phys.* **92**, 5397–5403 (1990).
- Trout, B. L. & Parrinello, M. Analysis of the dissociation of H_2O in water using first-principles molecular dynamics. *J. Phys. Chem. B* **103**, 7340–7345 (1999).
- Liu, Y. & Tuckerman, M. E. Protonic defects in hydrogen bonded liquids: Structure and dynamics in ammonia and comparison with water. *J. Phys. Chem. B* **105**, 6598–6610 (2001).
- Geissler, P. L., Dellago, C., Chandler, D., Hutter, J. & Parrinello, M. Autoionization in liquid water. *Science* **291**, 2121–2124 (2001).

Acknowledgements

M.E.T. was supported by the National Science Foundation (NSF) and Research Corporation.

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to M.E.T. (e-mail: mark.tuckerman@nyu.edu).

Nonlinear dynamics of ice-wedge networks and resulting sensitivity to severe cooling events

L. J. Plug* & B. T. Werner

Complex Systems Laboratory, Cecil and Ida Green Institute of Geophysics and Planetary Physics, University of California San Diego, La Jolla, California 92093-0225, USA

Patterns of subsurface wedges of ice that form along cooling-induced tension fractures, expressed at the ground surface by ridges or troughs spaced 10–30 m apart, are ubiquitous in polar lowlands¹. Fossilized ice wedges, which are widespread at lower latitudes, have been used to infer the duration^{2–4} and mean temperature^{5,6} of cold periods within Proterozoic² and Quaternary climates^{3–13}, and recent climate trends have been inferred from fracture frequency in active ice wedges¹⁴. Here we present simulations from a numerical model for the evolution of ice-wedge networks over a range of climate scenarios, based on the interactions between thermal tensile stress, fracture and ice wedges. We find that short-lived periods of severe cooling permanently alter the spacing between ice wedges as well as their fracture frequency. This affects the rate at which the widths of ice wedges increase as well as the network's response to subsequent climate change. We conclude that wedge spacing and width in ice-wedge networks mainly reflect infrequent episodes of rapidly falling ground temperatures rather than mean conditions.

Ice wedges originate with fractures opened in perennially frozen ground by tensile stress induced by falling winter temperatures^{15–17}.

* Present address: Department of Earth Sciences, Dalhousie University, Halifax B3H 4R2, Canada.

Fractures partially fill with ice from blowing or thawing snow, and then close when the ground warms^{16,17}. Because ice in fracture cavities is generally weaker than frozen ground, fractures usually follow the same path from year to year^{3,16–18}. Over thousands of years, V-shaped wedges of ice 3–5 m deep and up to several metres wide develop along recurring fracture paths³. Upward deformation at the ground surface forms ridges with heights ranging from decimetres to 1 m, revealing the position of growing ice wedges below³.

Although these basic processes relating to a single ice wedge have been the focus of previous investigations^{15–18}, the development of network patterns over longer timescales and their relationship to varying climates remains uncertain. The nonlinearity of fracture initiation and propagation, the influence of open fractures on subsequent fracturing during a single winter, and the long-term memory of fracture patterns from past winters stored in ice wedges all constitute building blocks of a system that could exhibit complex dynamical behaviour diverging from that of a single ice wedge.

Reflecting the essentially two-dimensional nature of ice-wedge networks¹⁹, we model fractures and ice wedges using a two-dimensional array of cells representing the top approximately five metres of frozen ground. Each cell contains either frozen ground or ice, and can contain a segment of an open fracture. All cells initially are assigned unfractured frozen ground, and thereafter evolve through repeated application of an algorithm that corresponds to one year in the model (Methods). To represent the thermal tensile stress that gives rise to fractures in frozen ground, a pre-fracture tensile stress (the thermal stress without reduction by fractures) that is uniform across the model lattice is incremented from 0 MPa to a specified maximum value. Fracture segments locally reduce this tensile stress by an amount that is proportional to the inverse square of distance and is maximized along a line perpendicular to the segment^{15,19}. As thermal stress increases, fractures initiate in cells where local stress

exceeds strength, and propagate cell-by-cell in two opposing directions along directions determined by stress imposed by open fracture segments, parameterized heterogeneity, and the energetic cost of propagation (higher in frozen ground than in ice). At the end of a fracture episode, the width of ice wedges is increased in cells with open fractures, all open fractures are closed, and values of tensile strength and propagation threshold are set to those for ice in all cells that contain fractures. The effect of a growing ice wedge on the ground surface is modelled by deforming the surface over the wedge by a volume equal to the volume of ice added. The relaxation

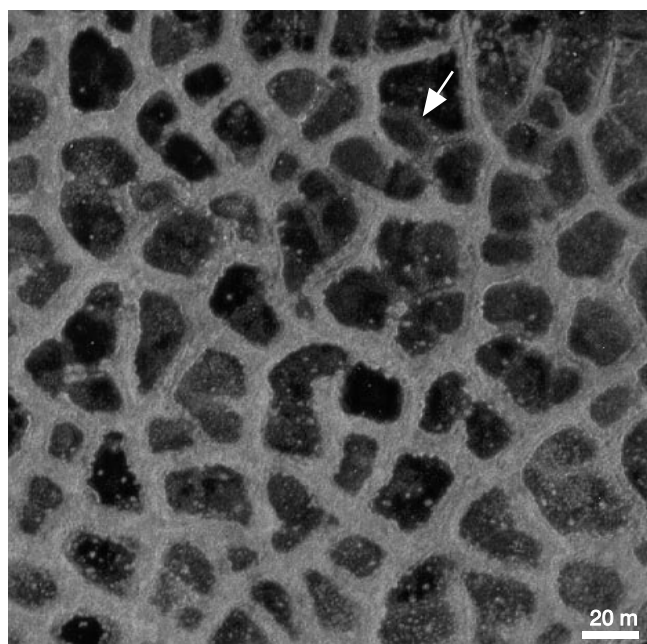


Figure 1 Near-infrared aerial photograph of an ice-wedge network on the floor of a drained lake near Espenberg, northwest Alaska. Ridges (0.25–1 m high, spaced 10–30 m apart) overlie ice wedges, and appear light because they are dry and covered by herbaceous vegetation. Troughs appear dark because of wet sedge-tundra vegetation or standing water. Arrow indicates low, barely visible ridge (see Fig. 2c and text). Photograph provided by the National Park Service.

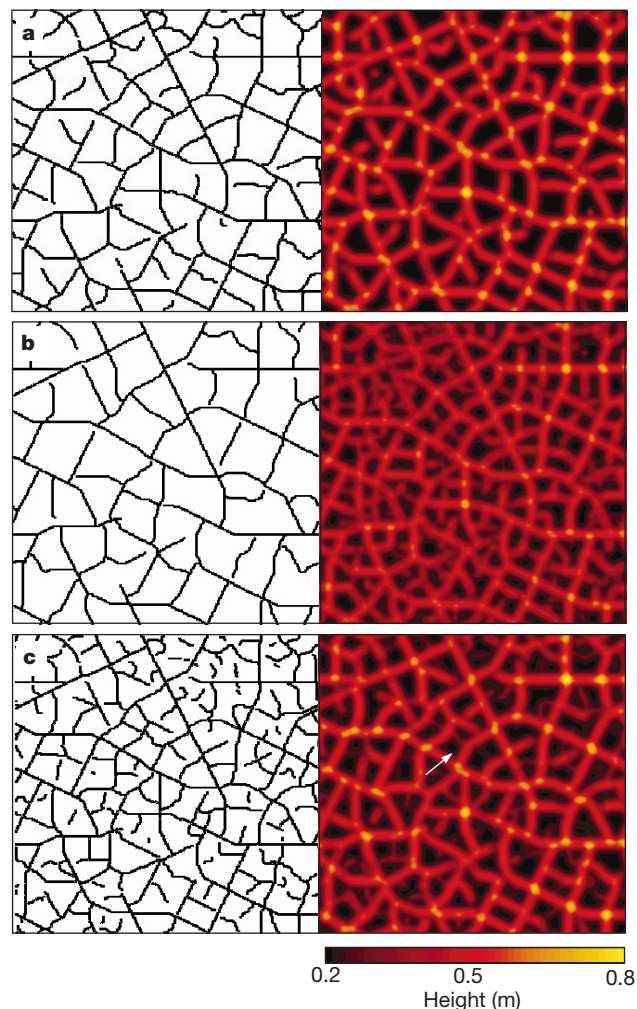


Figure 2 Fracture pattern after 1 yr, and surface elevation after 2,000 yr, for three model experiments. Left panels show fractures, right panels show height. The models are a uniform-stress network (a), a narrow-range network (b), and a lake-floor network (c). Arrow at bottom right indicates a low ridge developed over a modelled ice wedge that infrequently fractures. Relief below 0.2 m is shown black, representing shallow standing water common in many natural networks (see, for example, Fig. 1). All simulations use domains of 250×250 1-m² cells. To minimize variations between networks owing to random variations in first fractures rather than stress history, all simulations use the same random seed; hence, paths of long initial fractures are similar. Distributions of wedge spacing and orientation from the uniform-stress model at 2,000 yr (defined by a threshold ridge height of 0.3 m) match similarly sized regions of natural networks at Espenberg, Alaska, as evaluated with a two-dimensional Kolmogorov–Smirnov test¹⁹. Unlike most natural networks, modelled networks exhibit increased height (lighter patches) over some intersections between ice wedges. Possibly, natural ice-wedge growth increments near intersections are smaller because of nonlinear tensile stress reduction in the complicated region of an intersection, an effect not included in the model.

of morphology by soil creep is modelled with linear diffusion. The values of parameters were chosen to be consistent with existing measurements (Methods), many with broad uncertainties. Model results are not highly sensitive to parameter values¹⁹.

Our goal is to model the development of ice-wedge networks over thousands of years. Details pertaining to the shorter-timescale dynamics of fracture propagation, tensile stress field and frozen ground deformation are simplified or omitted, on the basis of the hypothesis that dynamics of these processes dissipate on the longer timescales of the pattern of interest^{20,21}, the ice-wedge network. This assumption is supported by comparisons of modelled to natural ice-wedge networks (Figs 1 and 2). In both modelled and natural networks, ice wedges range from long and sinuous to short and straight. Many ice wedges terminate at orthogonal intersections with other wedges, but some paths cross, forming four-way junctions. The spacing (15–20 m) and relative orientation (Fig. 3) of surface ridges and the mean width of ice wedges after 2,000 yr (1–3 m) are consistent with measured values from ice-wedge networks in northern Alaska and Canada^{3,19}. The similarity between natural networks and networks modelled using reasonable parameter values (Methods), and the relative insensitivity of model results to parameter values¹⁹, suggest that dynamics of modelled networks can be used to investigate natural networks, despite some parameters being poorly constrained by measurements (Methods). Qualitative model-based inferences are most robust.

The response of modelled networks to changes and variability in climate was investigated using four numerical experiments in which only the sequence of maximum annual stress was varied. In the first experiment, the severity of each winter's cooling episode was fixed by setting maximum tensile stress to the same value, 2 MPa, each year. The resultant ice-wedge network diverges from the initial fracture pattern (Fig. 2a), as illustrated by a decrease in spacing between ice wedges from 21.7 m in the first year to 19.2 m after 800 yr (Fig. 3a). Although most fractures are constrained to ice wedges, the divergence in spacing arises as different permutations in the order of ice-wedge fracturing permit the introduction of new fracture paths and hence new wedges. These new fractures open at

locations where tensile stress exceeds the strength of frozen ground before fractures that reduce the regional tensile stress open in nearby ice wedges. Switching the sequence of fracturing in two ice wedges can convert an orthogonal three-way (T-type) intersection into a crossing four-way intersection; in this case, the wedge that terminates at the T can be extended if it fractures before the perpendicular wedge. After approximately 800 yr, the modelled ice-wedge network attains a steady state, characterized by consistent spacing between fractures, ice wedges and surface ridges. 88% of ice wedges fracture each year, a frequency sufficient to form ridges more than 0.3 m high at the surface over all ice wedges. Long ice wedges tend to fracture more frequently because they contain more potential sites for fracture initiation, and because a fracture, once initiated in a wedge, generally propagates along its entire length. Open fractures in these long wedges reduce tensile stress in nearby short wedges, lowering their fracture frequency. The width of ice wedges reaches 2.8 ± 0.7 m after 2,000 yr, the high standard deviation reflecting the variation in frequency with which ice wedges fracture and the addition of new wedges during the first 800 yr. The width and frequency of fracture of ice wedges decrease as the ratio of ice strength to ground strength increases in the model; we expect that the values provided here are upper limits.

To investigate the sensitivity of ice-wedge width and spacing to climate variability, two experiments with maximum annual stress normally distributed around a mean of 2 MPa were performed, with standard deviation (s.d.) = 1 MPa (narrow-range) and s.d. = 1.75 MPa (broad-range). The pattern of ice wedges in the narrow-range network reaches an approximate steady state after about 1,200 yr, with a spacing between ice wedges of 13.1 m (Fig. 3b), whereas the broad-range network only attains approximate steady state after 1,800 yr, with ice wedges separated by 12.0 m (Fig. 3c). The fraction of ice wedges that fracture in a given year is reduced from, and more variable than, the fraction for the uniform-stress network: $57 \pm 15\%$ for the narrow-range network, and $48 \pm 23\%$ for the broad-range network. Spacing between surface ridges after 2,000 yr is 15.7 m for the narrow-range network and 16.8 for the broad-range network (Fig. 3b, c). These spacing values are intermediate between mean annual fracture spacing and ice-wedge spacing, because not all wedges added during high-stress winters fracture with sufficient frequency during subsequent winters to form ridges. Counter to intuition, ridge spacing in the broad-range network is greater than ridge spacing in the narrow-range network, despite reduced ice-wedge spacing. This reflects greater over-saturation by ice wedges in the broad-range network; therefore, fewer ice wedges fracture with sufficient frequency to form ridges. The patterns of surface ridges of both broad- and narrow-range networks achieve steady state more than 600 yr after the corresponding ice-wedge patterns; this is because ice wedges added during high-stress years are expressed at the surface only after many episodes of refreezing.

The sensitivity of ice-wedge networks to conditions at their time of initiation is illustrated in an experiment in which maximum tensile stress was set to 4 MPa for the first 4 yr of network development and to 0.8 MPa (below the fracture initiation threshold in frozen ground) thereafter. This scenario corresponds to network development following draining of a thaw lake²², in which minimal insulating snow accumulates in the wind-swept basin during the first winters after desiccation, but accumulation increases to approximately 0.5 m after 5 yr as snow-trapping plants colonize the basin¹⁶. Spacing of modelled wedges reaches 16.3 m after 4 yr (Figs 2c, 3d), and remains constant thereafter because the maximum tensile stress is below frozen ground strength. Despite developing principally under a maximum stress insufficient to fracture frozen ground, ice wedges in the lake-floor network grow at a rate comparable to that in other numerical experiments. More so than in the other experiments, ice wedges that are long, and therefore contain more potential sites for fracture initiation, tend to fracture

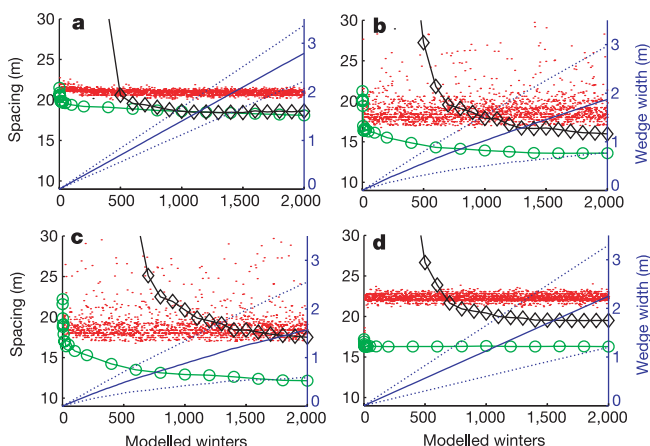


Figure 3 Properties of ice-wedge and fracture patterns versus time for four modelled networks. **a**, Uniform-stress network; **b**, narrow-range network; **c**, broad-range network; and **d**, lake-floor network. Green circles, spacing between ice wedges; red dots, spacing between fractures in the annual fracture pattern; black diamonds, spacing between surface ridges that form over wedges; blue lines, width of ice wedges (solid line, mean; dotted line, 1 s.d. envelope). Spacing is the median of 1,000 measurements of distance between fractures or ice wedges, measured along sample lines that cross a network at randomly selected angles and originate from randomly selected locations¹⁹. Spacing of surface ridges is measured between centre-lines of ridges at the 0.30-m contour interval, a height that approximates the characteristic roughness of tussock tundra surfaces.

with greater frequency than short wedges. Short ice wedges that rarely fracture do not appear (or barely appear) in the visual pattern of surface ridges because of relaxation of ground surface morphology (Fig. 2c). The spacing between surface ridges, 19.4 m, is consequently 20% greater than ice-wedge spacing (Fig. 3d).

We interpret these results as having the following implications. First, metre-wide fossil wedges indicate the existence (for at least thousands of years) of permafrost and conditions sufficient to fracture wedge ice; but they do not necessarily indicate sustained conditions sufficiently harsh to fracture frozen ground, because a brief exposure to high-tensile-stress conditions can lead to the formation of an ice-wedge network that then continues to develop under conditions insufficient for initial formation. Because networks are sensitive to initial and extreme conditions, the geographical distribution of networks might not correlate with mean climate, as has been suggested for modern¹⁰ and fossil^{2–10} networks. Second, wedge spacing⁷ mainly reflects infrequent severe winter conditions (periods of rapidly falling air temperature at low absolute temperature) during the formation of the network, because episodes of high maximum tensile stress add new wedges to a network. Third, measurements of ice-wedge width^{2–4} might provide minimum limits to the duration of cold climates, but the maximum duration is uncertain because growth rates are reduced by climatic variability and are widely variable between ice wedges. Long wedges are the most reliable measure of network duration because they form early in network development. Narrow fossil wedges are consistent with a cold, inherently variable climate, or with a moderate climate marginally suitable for ice-wedge development^{5,6}. Fourth, low rates of fracture in networks^{4,18} are consistent with a stationary but inherently variable climate, as well as with other climate scenarios, and so alone cannot be used to infer climate change. However, changes in fracture frequency across a population of ice wedges¹⁴, sustained over many winters, are consistent with climate change.

The characteristics of ice-wedge networks reflect an interplay between fracturing, ice-wedge growth and network development, all nonlinear processes with differing intrinsic timescales. Our model for the evolution of ice-wedge networks shows complex behaviour under a range of climate scenarios. This suggests the need for caution in inferring the response of ice-wedge networks (and that of other, more complicated, geomorphic systems) to climate change using simple models or conceptions of their behaviour. □

Methods

Stress

Thermal tensile stress in frozen ground is a complicated function of material-dependent rheology, absolute temperature and rate of cooling, and is highly uncertain in field settings. The maximum annual stress used for the uniform stress model, 2 MPa, is based on a viscoelastic rheological model for frozen ground before fracture¹⁵, a cooling rate of 10°C d^{-1} and a quasi-viscous parameter, $\bar{n}$, of $8 \times 10^{25} \text{ N}^3 \text{ m}^{-6} \text{ s}$, where $\bar{n}$ relates the rate of permanent quasi-viscous strain $\dot{\epsilon}$ under a constant sustained stress P in $P^3 = 2\bar{n}\dot{\epsilon}$ (ref. 15). An increase of maximum annual stress to 3.75 MPa, used as the +1 s.d. limit in the broad-range network, might be caused by a decrease in the absolute ground temperature from -10°C to -30°C , with a corresponding increase in $\bar{n}$ to $15 \times 10^{25} \text{ N}^3 \text{ m}^{-6} \text{ s}$ (ref. 15). Changes in absolute ground temperature are one route to changes in maximum annual tensile stress, and changes in the rate of temperature change are another¹⁵.

Local tensile stress, $\sigma_{kl}(\theta)$, the stress perpendicular to angle θ in cell kl , is taken to be pre-fracture thermal tensile stress, σ_T , minus the sum of reductions from nearby open fracture segments in cells ij , according to

$$\sigma_{kl}(\theta) = \sigma_T \left(1 - \sum_{ij} \frac{C}{(d_{ijkl})^n} \cos(\beta_{ijkl}) \sin(\alpha_{ijkl}) \right) \quad (1)$$

where d_{ijkl} is the distance (in m) from the centre of cell ij to the centre of cell kl , β_{ijkl} is the difference between the angle of the fracture in cell ij and angle θ , and α_{ijkl} is the angle between the line segment connecting cell kl and cell ij and the normal to the fracture segment in cell ij (ref. 19). Parameters are set to $n = 2$ and $C = 3 \text{ m}^2$ so that far-field modelled stress approximates the solution to two-dimensional elastic equations for stress around an infinitely long, straight fracture 4 m deep^{15,19}.

Fracture

Fractures are initiated where local tensile stress exceeds strength (set to 1 MPa for frozen

ground and 0.3 MPa for ice wedges, values chosen from the small range of laboratory measurements of small samples of ice and frozen silt^{19,23,24}, because *in situ* measurements of wedge ice and heterogeneous soil are lacking). Once initiated, fractures continue to propagate if the elastic strain energy released by opening of the fracture, G , exceeds the energy required to create new fracture surfaces, G_o , set to 1.5 J m^{-2} in ice-wedge cells²⁵ and to 8 J m^{-2} in frozen ground cells¹⁹. Both values have large uncertainties; their ratio determines the degree to which fractures tend to follow wedges. G is calculated from the pre-fracture tensile stress along the 5 m of fracture path preceding the position of the fracture tip, linearly weighted toward stresses near the tip, and from modelled shear stress across the fracture plane at the fracture tip¹⁹. The angle of fracture propagation is re-evaluated as a fracture tip enters a new cell, with $P(\theta, \Delta\theta)$, the probability that propagation angle changes from θ to $\theta + \Delta\theta$, given by a Boltzmann factor $\exp[\Delta G(\theta + \Delta\theta)/G_{\text{random}}]$. Here $\Delta G(\theta + \Delta\theta)$ is the change in net energy ($G - G_o$) released as a function of change in propagation angle and G_{random} parameterizes the magnitude of heterogeneity in frozen ground and in tensile stress. G_{random} is set to 0.5 J m^{-2} so that isolated fractures in the model follow irregular paths similar to paths of first fractures in frozen ground^{16,19}.

Ice-wedge and surface dynamics

The width of ice wedges is increased in cells with open fractures by a positive amount drawn from a normal distribution with both mean and standard deviation 0.0015 m (ref. 3). The effect of a growing ice wedge on the ground surface is modelled by raising the surface over the wedge by a volume equal to the volume of ice added, with the depth of fractures assumed to be 4 m. This elevation increment has a gaussian shape centred on the axis of the ice wedge, with standard deviation equal to the wedge width². This approach assumes the common natural case of a ridge over an ice wedge. Other modes of deformation can result in different morphologic responses¹; however, the cause of differences between them is poorly understood and is not the subject of our investigation. Measurements of morphologic relaxation in permafrost are lacking; we model relaxation as diffusion²⁶ with diffusion constant $5 \times 10^{-3} \text{ m}^2 \text{ yr}^{-1}$.

Received 19 December 2001; accepted 26 April 2002; doi:10.1038/nature00796.

- French, H. M. *The Periglacial Environment* (Addison Wesley, Harlow, 1996).
- Williams, G. E. et al. in *Earth's Glacial Record* (eds Deynoux, M. et al.) 146–164 (Cambridge Univ. Press, Cambridge, 1994).
- Black, R. F. in *Glacial Geomorphology* (ed. Coates, D.) 247–275 (State Univ. New York, Binghamton, 1974).
- Johnson, W. H. Ice-wedge casts and relic patterned ground in central Illinois and their environmental significance. *Quat. Res.* **33**, 51–72 (1990).
- Renssen, H. & Isarin, R. F. B. The two major warming phases of the last deglaciation at ~ 14.7 and $\sim 11.5 \text{ ka}$ BP in Europe: climate reconstructions and AGCM experiments. *Glob. Planet. Change* **30**, 117–153 (2001).
- Isarin, R. F. B. Permafrost distribution and temperatures in Europe during the Younger Dryas. *Permafrost Periglac. Process* **8**, 313–333 (1997).
- Maizels, J. K. Frequency of relic frost-fissure structures and prediction of polygon pattern: A quantitative approach. *Biul. Peryglac.* **30**, 67–89 (1986).
- Walters, J. C. Ice-wedge casts and relic polygonal patterned ground in North-East Iowa, USA. *Permafrost Periglac. Process* **5**, 269–282 (1994).
- Owen, L. A. et al. Relic permafrost structures in the Gobi of Mongolia: age and significance. *J. Quat. Sci.* **13**, 539–547 (1998).
- Pewe, T. L. *Proc 1st Int. Permafrost Conf.* 76–81 (Publication 1287, National Academy of Science and National Research Council, Washington DC, 1966).
- Black, R. F. Periglacial features indicative of permafrost: Ice and soil wedges. *Quat. Res.* **6**, 3–26 (1976).
- Borns, H. W. Late glacial ice-wedge casts in northern Nova Scotia, Canada. *Science* **148**, 1223–1226 (1965).
- Grosso, S. A. & Corte, A. E. Pleistocene ice wedge casts at 34 degrees S; Eastern Andes piedmont, south-west of South America. *Geogr. Ann.* **71**, 125–136 (1989).
- Kasper, J. N. & Allard, M. Late-Holocene climatic changes as detected by the growth and decay of ice wedges on the southern shore of Hudson Strait, northern Québec, Canada. *Holocene* **11**, 563–577 (2001).
- Lachenbruch, A. H. Mechanics of thermal contraction cracks and ice-wedge polygons in permafrost. *Spec. Pap. Geol. Soc. Am.* **70** (1962).
- Mackay, J. R. The first seven years (1978–1985) of ice-wedge growth; Illisarvik experimental drained lake site, western Arctic coast. *Can. J. Earth Sci.* **23**, 1782–1795 (1986).
- Mackay, J. R. Air temperature, snow cover, creep of frozen ground, and the time of ice-wedge cracking, western Arctic coast. *Can. J. Earth Sci.* **30**, 1720–1729 (1993).
- Mackay, J. R. The frequency of ice-wedge cracking (1967–1987) at Garry Island, western Arctic coast, Canada. *Can. J. Earth Sci.* **29**, 236–248 (1992).
- Plug, L. J. & Werner, B. T. Fracture networks in frozen ground. *J. Geophys. Res.* **106**, 8599–8613 (2001).
- Werner, B. T. Complexity in natural landform patterns. *Science* **284**, 102–104 (1999).
- Nicolas, G. & Prigogine, I. *Self-organization in Nonequilibrium Systems* (Wiley, New York, 1977).
- Hopkins, D. M. Thaw lakes and thaw sinks in the Imuruk Lake area, Seward Peninsula Alaska. *J. Geol.* **57**, 119–131 (1949).
- Zhu, Y. & Carbee, D. L. *Tensile Strength of Frozen Silt* Report 87-14 (Army Cold Regions Research and Engineering Laboratory, Hannover, New Hampshire, 1987).
- Gold, L. W. Engineering properties of freshwater ice. *J. Glaciol.* **19**, 197–212 (1977).
- Goodman, D. J. & Tabor, D. Fracture toughness of ice: An account of some new experiments. *J. Glaciol.* **21**, 651–660 (1978).
- Hanks, T. C., Bucknam, R. C., Lajoie, K. R. & Wallace, R. E. Modification of wave-cut and faulting controlled landforms. *J. Geophys. Res.* **89**, 5771–5790 (1984).

Acknowledgements

We thank R. Anderson and B. Hallet for reviews, and the late D.M. Hopkins for comments on an earlier version. This work was supported by the NSF, Arctic Natural Sciences

Program, the Andrew W. Mellon Foundation, and the National Park Service, Bering Land Bridge National Preserve.

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to L.P.
(e-mail: lplug@ucsd.edu).

Waning buoyancy in the crustal roots of old mountains

Karen M. Fischer

Department of Geological Sciences, Brown University, Providence, Rhode Island 02912, USA

When mountains form through the collision of lithospheric plates, uplift of the Earth's surface is accompanied by thickening of the crust, and the buoyancy of these deep crustal roots (relative to the surrounding mantle) is thought to contribute to the support of mountain topography. Once active tectonism ceases, continuing erosion will progressively wear away surface relief. Here I provide new constraints on how crustal roots respond to erosional unloading over very long timescales. In old collisional mountain belts, ratios of surface relief to the thickness of the underlying crustal root are observed to be smaller than in young mountains. On the basis of gravity data, this trend is best explained by a decrease in the buoyancy of the crustal root with greater age since the most recent mountain-building episode—which is consistent with metamorphic reactions^{1,2} produced by long-term cooling. An approximate balance between mountain and root mass anomalies suggests that the continental lithosphere remains weak enough to permit exhumation of crustal roots in response to surface erosion for hundreds of millions of years. The amount of such uplift, however, appears to be significantly reduced by progressive loss of root buoyancy.

Processes such as lithospheric delamination and rifting may strip away the crustal roots of some collisional mountain regions, but substantial crustal roots have survived in many mountain belts over hundreds of millions to billions of years^{3–13}. Unless the lithosphere is mechanically very rigid, post-tectonic erosion of mass from the surface should be accompanied by some uplift of a buoyant crustal root and inflow of mantle. If the lithosphere is very weak, the region should be in local isostatic equilibrium, and the net change in mass over time would be zero. To examine how mountain crustal roots evolve over time, surface relief, crustal thickness and gravity data were compared in young collisional mountain belts and in old orogens where some crustal root is preserved. Crustal thickness was constrained seismically, primarily by refraction and reflection studies but in a few cases by teleseismic receiver function and surface wave analyses^{3–22}.

Ratios of surface relief (h) to crustal root magnitude (m) reveal a pronounced secular trend (Figs 1 and 2). In active or very young collisional orogens such as the Alps, Carpathians, Tien Shan and the Tibetan Plateau^{17–22}, these ratios (R) span a wide range, but are systematically higher than in older orogenic belts, particularly those where contraction ceased more than 100 million years (Myr) ago (Fig. 1a). Substantial zones of thickened crust are observed beneath some Proterozoic age orogens^{10–13}. Little to no topography is associated with these apparent crustal roots, and thus R values for these regions lie close to zero. The decrease in R with age becomes

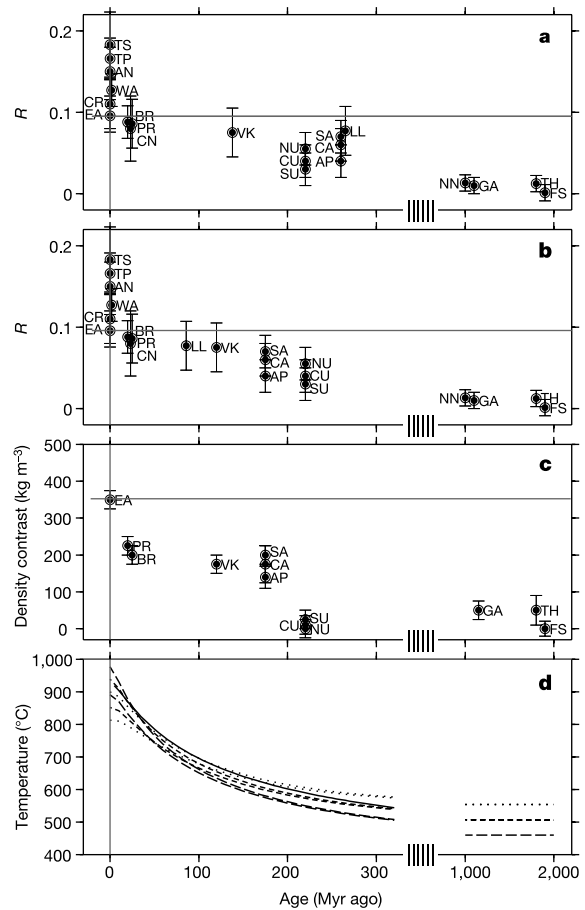


Figure 1 Comparison of surface topography to crustal root thickness, crustal root buoyancy, and crustal root temperature for young and old collisional mountain belts. **a**, Ratios (R) of mountain surface relief (h) to crustal root thickness (m) as a function of the time since collision ceased. **b**, R values as a function of time since the last major tectonic event in the region. **c**, Differences between crustal root and mantle density that best fit observed Bouguer gravity anomalies, assuming the seismically constrained shape of the crustal root and allowing a single value for upper crust density to vary. Error bars are 95% confidence limits. TP, Tibetan Plateau; TS, Tien Shan; AN, central Andes; WA, western Alps; CR, Carpathians; EA, eastern Alps; PR, central Pyrenees; BR, Brooks range; CN, Cantabrian mountains; VK, Verkhoyansk mountains; NU, northern Urals; CU, central Urals; SU, southern Urals; SA, southern Appalachians; CA, central Appalachians; AP, Appalachian Plateau; LL, Lachlan Orogen; NN, Namaqua–Natal Orogen; GA, northwest Grenville Orogen; TH, Trans-Hudson Orogen; FS, Svecofennian Orogen. **d**, Temperature at the Moho for a 45-km-thick crust from analytical and finite difference cooling models. Analytical calculations cool from a geotherm parameterized by initial surface heat flow (q_0) and crustal heat production (A) to an infinite time geotherm with lower heat flow (q_f) but the same A value; q_f of 50 mW m^{-2} and A of $0.8 \mu\text{W m}^{-3}$ (dotted lines); q_f of 45 mW m^{-2} and A of $0.7 \mu\text{W m}^{-3}$ (short dashed lines); q_f of 40 mW m^{-2} and A of $0.6 \mu\text{W m}^{-3}$ (long dashed lines). In each case q_0 values of 70 mW m^{-2} and 65 mW m^{-2} correspond to the hotter and cooler initial geotherm, respectively. Initial thermal lithospheric thicknesses are 60–90 km and final thicknesses are 190–230 km. In the one-dimensional finite difference calculation (solid line) a starting geotherm (q_0 of 70 mW m^{-2} and A of $0.7 \mu\text{W m}^{-3}$) over a half-space was allowed to cool freely given continuous constant crustal heat production. For cooling since 320 Myr ago or later, finite difference cooling rates are comparable to the analytical calculation with the same q_0 and A , but slightly slower. Over longer times, the finite difference model continues to cool to unreasonably low heat flow and large lithospheric thickness, and is not shown. Conductivity in all cases is $2.6 \text{ W m}^{-1} \text{ }^{\circ}\text{C}^{-1}$, and mantle potential temperature is $1,300 \text{ }^{\circ}\text{C}$ with an adiabatic gradient of $0.3 \text{ }^{\circ}\text{C km}^{-1}$. These calculations are not meant to replicate the temperature history of specific orogens. Rather, they are intended to show that significant cooling is possible over 200–300-Myr timescales.

even more pronounced when the data are plotted as a function of the last major thermotectonic event in the region (Fig. 1b). Some of the old mountain belts bordered regions where the lithosphere underwent post-collisional extension and rifting (the Appalachians, Verkhoyansk and Lachlan mountains) and hence appear in Fig. 1b at younger ages.

To evaluate possible contributions to the observed R values from mantle density, density variations across the various mountain belts in the upper 200 km of the mantle were estimated from a global shear-wave velocity model (S. Grand, personal communication). Some of the younger orogens have positive velocity anomalies at a lateral scale comparable to the mountains that suggest the presence of cold subducting lithosphere, a finding consistent with other tomographic studies. If R values for these regions were corrected for the density anomalies implied by these velocities, they would be shifted to higher values, enhancing the observed decrease in R with age. This result suggests that some of the apparent spread in R values for young orogens may reflect variations in lithospheric processes such as delamination. Other possible factors include different lower crustal compositions, temperatures, and possibly degrees of partial melt^{1,2,23}. Very little variation in mantle velocity is correlated with the older crustal roots, and mantle buoyancy cannot explain the overall trend in observed R values.

Assuming that old and young orogens reflect common mountain-building processes, two end-member hypotheses are consistent with a decrease in R with time after orogenesis or adjacent continental rifting (Fig. 2). For the first hypothesis, if the rigidity of the lithosphere increased with time as the mountains were eroded, the ageing lithosphere could partially resist isostatic uplift of the crust, maintaining a thick, very buoyant crustal root. For example, if a buoyant mass anomaly is imbedded in an elastic lithosphere close to the Moho, upward deflection of the surface as a result of this load will be diminished at small spatial scales but permitted at long wavelengths, with the transition between these responses governed by elastic plate thickness²⁴. The old crustal roots

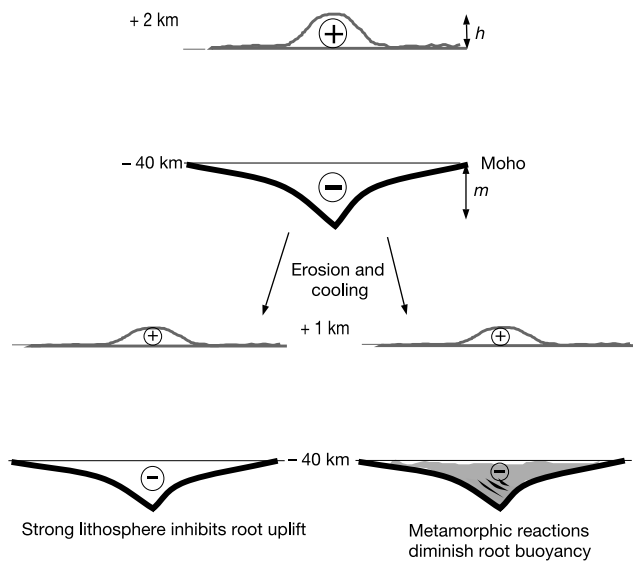


Figure 2 Schematic view of two hypotheses of how $R (= h/m)$ may decrease from young (upper) to old (lower) orogens. If lithospheric plate strength increases rapidly, it may inhibit crustal root uplift resulting in a thick, very buoyant root (black line in lower left). If the density of the crustal root relative to the mantle increases with time, perhaps as a result of metamorphic transformation (shading in crustal root, lower right), a thick crustal root with buoyancy close to isostatic equilibrium could be maintained (lower right). Gravity data are more consistent with the second hypothesis. One possible composition for old crustal roots is a mix of granulite (light grey shading, lower right) and localized zones of eclogite (darker bands in crustal root, lower right).

in this study correspond to dominant wavelengths of roughly 300 km to 600 km. However, given existing debate on elastic plate thicknesses for these regions^{24–27}, inhibition of their uplift is possible but cannot be precisely estimated. For the second hypothesis, if the density contrast between the crustal root and the mantle decreased with time, a thick crustal root and a mass balance at or close to isostatic equilibrium could be maintained. Some uplift of the crustal root could occur in response to surface erosion, but the amount of uplift would be reduced by the progressive loss of crustal root buoyancy.

The first hypothesis predicts much larger negative Bouguer gravity anomalies associated with crustal roots than does the second, allowing clear testing of these end-member scenarios. For 13 mountain belts, observed Bouguer anomaly profiles^{15,25–29} were

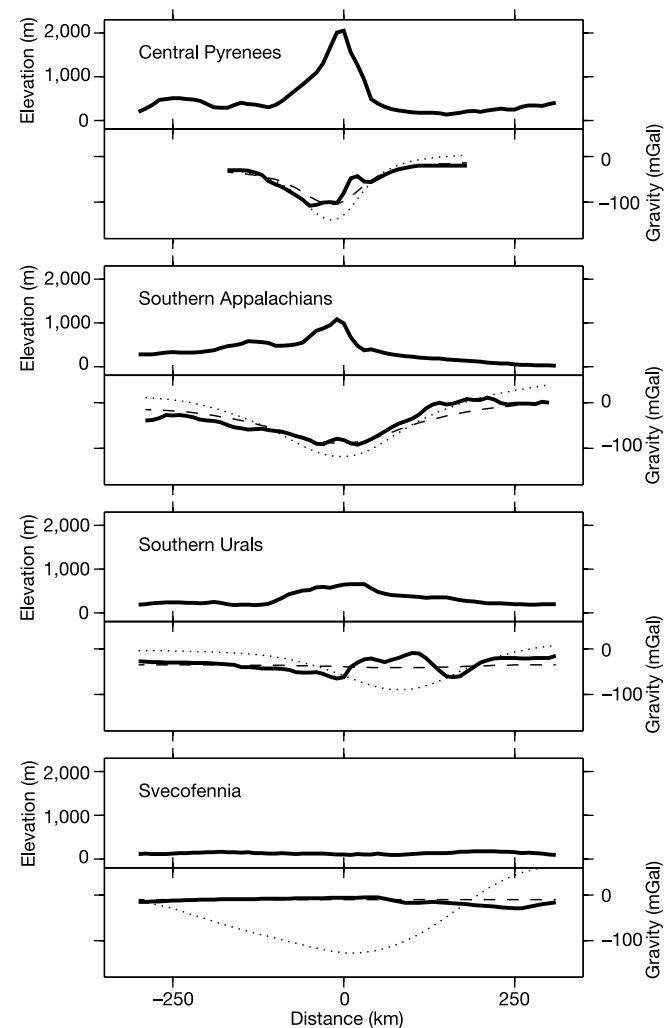


Figure 3 Observed topography and observed and predicted gravity profiles across four mountain belts of increasing thermotectonic age. Topography profiles (solid line in upper panel for each region) are from the TerrainBase 5-min Global Elevation Database and were constructed by averaging six individual profiles oriented normal to the strike of the mountains that collectively span 120 km along strike. Observed Bouguer gravity^{15,27–29} (solid line in lower panel for each region) is matched well by gravity predicted for the best-fitting crustal root and upper crust density values (dashed line). Gravity for a 350 kg m⁻³ root buoyancy (dotted line) is shown for reference and in all cases it overpredicts the amplitude of the observed Bouguer gravity anomaly at 95% confidence. The Proterozoic Svecofennian orogen offers a particularly striking example of decreased root buoyancy with large thermotectonic age. There a crustal root of at least 15 km in thickness¹⁰ is accompanied by virtually no variation in observed surface elevation or Bouguer gravity²⁷. The net buoyancy of this crustal root must therefore be close to zero.

compared to gravity predicted by two-dimensional density models which assumed the seismologically constrained depth and dimensions of the crustal root. Density within the crustal root and density in a uniform crustal layer above the root were treated as two independent parameters and were allowed to vary until best-fitting values which maximized gravity variance reduction were found.

The gravity data resolve a significant decline in root buoyancy with increasing thermotectonic age (Fig. 1c), consistent with the second hypothesis. The eastern Alps, the active collision zone with the smallest R value, yields a best-fitting crustal root density anomaly of 350 kg m^{-3} , and similar modelling for other young orogenic regions, such as the Tibetan Plateau, indicates even greater density contrasts between crustal root and mantle. In contrast, crustal root buoyancy is significantly smaller for all orogens whose last major thermotectonic event was 20 Myr ago or earlier. Profiles illustrating gravity modelling in older mountain belts (Fig. 3) show that the best-fitting density values (dashed lines) do in fact match the observed Bouguer gravity anomalies (solid lines) much better than models with the 350 kg m^{-3} reference root buoyancy from the eastern Alps (dotted lines).

The philosophy of this modelling approach is to provide a consistent comparison of average density and overall buoyancy between crustal roots from different regions. Because lateral variations in crustal density probably occur, it is not surprising that these simple models do not capture some shorter wavelength variations in the observed Bouguer anomalies. However, in the southern Urals and the Pyrenees, gravity modelling that allows for lateral density variations associated with sedimentary basins and other features still concludes that the lower crust is unusually dense^{15,28}.

The average crustal root densities that best fit the old mountain belt gravity data overlap with the average densities required to isostatically compensate for surface relief at 95% confidence, with the exception of the central and northern Urals. This correlation strongly suggests that exhumation of the root in response to surface erosion continues to occur over timescales of hundreds of millions of years. In some regions, maximum root thickness is clearly laterally displaced from peak surface relief²⁻⁷, as shown for the southern Urals in Fig. 3 by the horizontal offset between maximum surface elevation and the most negative predicted Bouguer gravity anomaly values. This type of observation may indicate that some of the surface load is regionally supported by the mechanical lithosphere, as has been suggested for both young and old orogens²⁴⁻²⁷, although local isostatic equilibrium cannot be ruled out owing to the possibility of lateral variations in density. In either case, however, the lithosphere must be weak enough to permit some root uplift.

Metamorphic phase transitions in the lower crust induced by cooling provide a plausible explanation for the apparent decrease in crustal root buoyancy with time. Finite difference (S. Zaranek, personal communication) and analytical calculations illustrate that significant cooling at lower crustal depths may persist over timescales of 200–300 Myr (Fig. 1d). The analytical models (dotted and dashed lines) describe cooling from initial geotherms with surface heat flow values of 65 to 70 mW m^{-2} to geotherms with heat flow values of 40 to 50 mW m^{-2} at infinite time, given constant crustal heat production values of 0.6 to $0.8 \mu\text{W m}^{-3}$. In the one-dimensional finite difference calculation cooling occurs freely from an initial heat flow of 70 mW m^{-2} with continuous heat production of $0.7 \mu\text{W m}^{-3}$ (solid line), and the rate of cooling over the first 320 Myr is only slightly slower than in the corresponding analytical model. The highest initial Moho temperatures are consistent with xenolith data from the Tibetan Plateau²³, while the lower initial temperatures may be appropriate for other active orogens².

If young crustal roots are on average mafic granulite^{30,31}, cooling over these temperature ranges could produce an increase in lower crustal density with respect to the mantle of 100 – 150 kg m^{-3} within

the granulite facies alone, largely because of an increase in the volume fraction of garnet^{1,2}. Further loss of buoyancy could occur if the deepest portions of the root cooled into the eclogite field. Pure mafic eclogite is denser than the mantle and with its mantle-like seismic velocities^{30,31} would probably be imaged below the seismically observed Moho. However, if transformation to eclogite occurred only locally where the presence of water catalysed the retrograde granulite–eclogite reaction, as has been observed in cracks and shear zones in a number of exposed crustal sections³², the resulting granulite–eclogite mix could lie above the seismic Moho with a neutrally or slightly buoyant density (Fig. 2, lower right). Variation to intermediate (more silica-rich) lower crustal compositions in certain regions³⁰ may also explain some of the range in apparent root density. For example, pure andesite eclogite would be close to neutrally buoyant and an andesite granulite–eclogite mix would be less dense than the mantle². In principle, retrograde metamorphism could increase the density of the deepest continental crust (deeper than roughly 40 km) in any region that experiences cooling over comparable temperatures and timescales. In the crustal layer above mountain roots, retrograde metamorphism may increase or at some conditions decrease crustal density^{2,32}. However, any laterally uniform density changes in this layer would not contribute to the observed secular decrease in R .

This study, based on global observations, indicates that erosion of mountains at the surface is accompanied by exhumation of their crustal roots for hundreds of millions of years after active tectonism has ceased, but that loss of crustal root buoyancy, possibly produced by retrograde metamorphism, reduces the amount of root uplift. In Proterozoic orogens, net crustal root buoyancy lies close to zero. Given that little cooling occurs past 300 Myr (Fig. 1d), the simplest explanation for this observation is that any more buoyant roots of similar age were erased by exhumation as the last remnants of mountain topography were eroded away. □

Received 18 February; accepted 21 May 2002; doi:10.1038/nature00855.

- Jull, M. & Kelemen, P. B. On the conditions for lower crustal convective instability. *J. Geophys. Res.* **106**, 6423–6446 (2001).
- Bousquet, R., Goffé, B., Henry, P., Le Pichon, X. & Chopin, C. Kinematic, thermal and petrological model of the Central Alps: Lepontine metamorphism in the upper crust and eclogitisation of the lower crust. *Tectonophysics* **273**, 105–127 (1997).
- Carbonell, R. *et al.* Seismic wide-angle constraints on the crust of the southern Urals. *J. Geophys. Res.* **105**, 13755–13777 (2000).
- Knapp, J. H. *et al.* Seismic reflection fabrics of continental collision and post-orogenic extension in the Middle Urals, central Russia. *Tectonophysics* **288**, 115–126 (1998).
- Schuller, W., Morozov, I. B. & Smithson, S. B. Crustal and uppermost mantle velocity structure of Northern Eurasia along the Profile Quartz. *Bull. Seismol. Soc. Am.* **87**, 414–426 (1997).
- Taylor, S. R. in *Geophysical Framework of the Continental United States Memoir 172* (eds Pakiser, L. C. & Mooney, W. D.) 317–348 (Geological Society of America, Boulder, CO, 1989).
- Li, A., Fischer, K. M., van der Lee, S. & Wyssession, M. E. Crust and upper mantle discontinuity structure beneath eastern North America. *J. Geophys. Res.* **107**, 10.1029/2001JB000190 (2002).
- Belousov, V. V. *et al.* Structure of the crust and upper mantle of the [former] USSR. *Int. Geol. Rev.* **34**, 345–444 (1992).
- Clitheroe, G., Gudmundsson, O. & Kennett, B. L. N. The crustal thickness of Australia. *J. Geophys. Res.* **105**, 13697–13713 (2000).
- GGT/SVEKA Working Group, Korsmann, K., Korja, T., Pajunen, M. & Virransalo, P. The GGT/SVEKA transect: Structure and evolution of the continental crust in the Paleoproterozoic Svecofennian orogen in Finland. *Int. Geol. Rev.* **41**, 287–333 (1999).
- Baird, D. J., Nelson, K. D., Knapp, J. H., Walters, J. J. & Brown, L. D. Crustal structure and evolution of the Trans-Hudson orogen: Results from seismic reflection profiling. *Tectonics* **15**, 416–426 (1996).
- Nguiri, T. K. *et al.* Crustal structure beneath southern Africa and its implications for the formation and evolution of the Kaapvaal and Zimbabwe cratons. *Geophys. Res. Lett.* **28**, 2501–2504 (2001).
- Rondenay, S., Bostock, M. G., Hearn, T. M., White, D. J. & Ellis, R. M. Lithospheric assembly and modification of the SE Canadian Shield: Abitibi–Grenville Teleseismic Experiment. *J. Geophys. Res.* **105**, 13735–13754 (2000).
- Wissinger, E. S., Levander, A. & Christensen, N. I. Seismic images of crustal duplexing and continental subduction in the Brooks Range. *J. Geophys. Res.* **102**, 20847–20871 (1997).
- Vacher, P. & Souriau, A. A three-dimensional model of the Pyrenean deep structure based on gravity modelling, seismic images and petrological constraints. *Geophys. J. Int.* **145**, 460–470 (2001).
- Fernández-Viejo, G. *et al.* Crustal transition between continental and oceanic domains along the North Iberian margin from wide angle seismic and gravity data. *Geophys. Res. Lett.* **25**, 4249–4252 (1998).
- Musacchio, G., Zappone, A., Cassinis, R. & Scarascia, S. Petrographic interpretation of a complex seismic crust–mantle transition in the central-eastern Alps. *Tectonophysics* **294**, 75–88 (1998).
- Marchant, R. H. & Stampfli, G. M. Subduction of continental crust in the Western Alps. *Tectonophysics* **269**, 217–235 (1997).
- Chalot-Prat, F. & Girdacea, R. Partial delamination of continental mantle lithosphere, uplift-related

- crust-mantle decoupling, volcanism and basin formation: a new model for the Pliocene-Quaternary evolution of the southern East-Carpathians, Romania. *Tectonophysics* **327**, 83–107 (2000).
20. James, D. E. & Sacks, I. S. in *Geology and Ore Deposits of the Central Andes* Spec. Pub. 7 (ed. Skinner, B. J.) 1–25 (Society of Economic Geologists, Littleton, Colorado, 1999).
 21. Owens, T. J. & Zandt, G. Implications of crustal property variations for models of Tibetan plateau evolution. *Nature* **387**, 37–43 (1997).
 22. Mahdi, H. & Pavlis, G. L. Velocity variations in the crust and upper mantle beneath the Tien Shan inferred from Rayleigh wave dispersion: Implications for tectonic and dynamic processes. *J. Geophys. Res.* **103**, 2693–2703 (1998).
 23. Hacker, B. R. *et al.* Hot and dry deep crustal xenoliths from Tibet. *Science* **287**, 2463–2466 (2000).
 24. Forsyth, D. W. Subsurface loading and estimates of the flexural rigidity of the continental lithosphere. *J. Geophys. Res.* **90**, 12623–12632 (1985).
 25. Stewart, J. & Watts, A. B. Gravity anomalies and spatial variations of flexural rigidity at mountain ranges. *J. Geophys. Res.* **102**, 5327–5352 (1997).
 26. McKenzie, D. & Fairhead, D. Estimates of the effective elastic thickness of the continental lithosphere from Bouguer and free air gravity anomalies. *J. Geophys. Res.* **102**, 27523–27552 (1997).
 27. Djomani, Y. H. P., Fairhead, J. D. & Griffin, W. L. The flexural rigidity of Fennoscandia: reflection of the tectonothermal age of the lithospheric mantle. *Earth Planet. Sci. Lett.* **174**, 139–154 (1999).
 28. Döring, J. & Götz, J.-J. The isostatic state of the southern Urals crust. *Geol. Rundsch.* **87**, 500–510 (1999).
 29. Hittelman, A. M., Kinsfather, J. O. & Meyers, H. *Geophysics of North America* [CD-ROM] (National Geophysical Data Center, Boulder, CO, 1990).
 30. Rudnick, R. L. & Fountain, D. M. Nature and composition of the continental crust: a lower crustal perspective. *Rev. Geophys.* **33**, 267–309 (1995).
 31. Christensen, N. I. & Mooney, W. D. Seismic velocity structure and composition of the continental crust: A global view. *J. Geophys. Res.* **100**, 9761–9788 (1995).
 32. Hacker, B. R. in *Subduction Top to Bottom* Geophysical Monograph 96 (eds Pakiser, L. C. & Mooney, W. D.) 337–346 (American Geophysical Union, Washington DC, 1996).

Acknowledgements

I thank S. Zaranek for the finite difference cooling calculations, S. Grand for the global shear-wave velocity model, G. Abers for his gravity code, and D. Forsyth, D. Scheirer and Y. Liang for discussions. This research was supported by the NSF Geophysics Program.

Competing interests statement

The author declares that she has no competing financial interests.

Correspondence and requests for materials should be addressed to the author (e-mail: Karen_Fischer@brown.edu).

Bird-like fossil footprints from the Late Triassic

Ricardo N. Melchor^{*†}, Silvina de Valais[‡] & Jorge F. Genise^{*‡}

^{*} CONICET & [†] Universidad Nacional de La Pampa, Av. Uruguay 151, 6300 Santa Rosa, La Pampa, Argentina

[‡] Museo Paleontológico “Egidio Feruglio”, Av. Fontana 140, 9100 Trelew, Chubut, Argentina

The study of fossilized footprints and tracks of dinosaurs and other vertebrates has provided insight into the origin, evolution and extinction of several major groups and their behaviour; it has also been an important complement to their body fossil record^{1–4}. The known history of birds starts in the Late Jurassic epoch (around 150 Myr ago) with the record of *Archaeopteryx*⁵, whereas the coelurosaurian ancestors of the birds date back to the Early Jurassic⁶. The hind limbs of Late Triassic epoch theropods lack osteological evidence for an avian reversed hallux and also display other functional differences from birds⁷. Previous references to suggested Late Triassic to Early Jurassic bird-like footprints have been reinterpreted as produced by non-avian dinosaurs having a high angle between digits II and IV^{8,9} and in all cases their avian affinities have been challenged¹⁰. Here we describe well-preserved and abundant footprints with clearly avian characters from a Late Triassic redbed sequence of Argentina^{11,12}, at least 55 Myr before the first known skeletal record of birds. These footprints document the activities, in an environment inter-

preted as small ponds associated with ephemeral rivers, of an unknown group of Late Triassic theropods having some avian characters.

Numerous small, bird-like footprints (Fig. 1) were discovered in, and collected from, the middle part of the Late Triassic Santo Domingo formation¹³ from northwest Argentina, La Rioja Province (28° 32' S, 68° 45' W). The age of the unit has been established on the basis of its fossil content and a radiometric date, and is further supported by lithologic comparison with the well-dated Late Triassic Los Colorados formation¹¹. The unit at the studied locality has produced remains of *Rhexoxylon*¹¹, a wood morphogenus only reported from Middle to Late Triassic rocks of Gondwana¹⁴. Concurrently, an interbedded basalt flow located about 80 m above the track-bearing horizons yielded an ⁴⁰Ar/³⁹Ar plateau age of 212.5 ± 7.0 Myr ago (step-heating analysis on albite crystal)¹², which suggests a Norian–Rhaetian age for the basalt according to recent calibrations of the Triassic–Jurassic timescale^{15–17}. The tracks occur in two trampled bedding planes (each up to 1.9 square metres) having a vertical spacing of less than 10 cm. Footprints appear with variable density and quality of preservation on the worked bedding planes (Fig. 1a); they are better preserved in the lower horizon, both as natural casts and moulds. In its central part (composed of claystone) they exhibit a high density (up to 520 footprints per square metre) and frequent overprinting: the footprints are moderately to poorly preserved and may be deeply impressed (up to 2.5 mm deep). There is a gradual transition in preservational quality towards the margins (composed of siltstone), where there are sparse sharp, shallower, well-preserved footprints having distinct pad impressions. This variation in footprint preservation suggests that the substrate in the central area of the trampled horizons was more watery than the marginal areas. Sedimentological analysis indicates that the track-bearing sediments were deposited in shallow ponds associated with an ephemeral fluvial system.

Table 1 summarizes the measurements of 50 distinct footprints, including eight tracks (a total of thirty footprints) and twenty isolated footprints (Fig. 1). The tracks are bipedal, displaying high pace angulation and a straight or slightly curved path; they lack a preferred relative orientation (Fig. 1a, b). The footprints frequently show a positive (inward) rotation in relation to the midline, the pace length being 2.4 to 4.5 times the mean footprint length (without hallux); they may be tetradactyl (58%) or tridactyl (42%) and, exclusive of the hallux impression, are consistently wider than their length. Tridactyl and tetradactyl footprints display a similar overall morphology and may be present in the same track; it is thus evident that they were left by the same producer, the hallux imprints not always being preserved. Digit impressions are slender (maximum

Table 1 Summary of measured track and footprint parameters

	Mean	Minimum	Maximum	n
Footprint length*	27.5	18.7	38.6	50
Footprint length†	36.3	29.4	46.3	29
Width	34.6	27.7	42.7	49
Length I	6.4	2	11.4	29
Length II	14.8	10.6	19.7	49
Length III	20.3	14.5	31.7	49
Length IV	17.3	10.9	22.9	48
Angle II–IV	115°	87°	137°	49
Angle II–III	59.5°	36°	74°	49
Angle III–IV	56.9°	35°	77°	49
Angle I–III	160°	133°	180°	28
Pace angulation	166.2°	107.5°	178°	15
Stride length	193.1	137	245	14
Departure midline	10.9°	0°	29°	29
Length/width	0.80	0.65	0.91	49
I/II‡	0.31	0.13	0.73	29
II/III‡	0.74	0.45	0.98	49
IV/III‡	0.87	0.49	1.31	48

All linear measurement in millimetres.

*Length without hallux. †Length including hallux. ‡Digit impression length ratios.

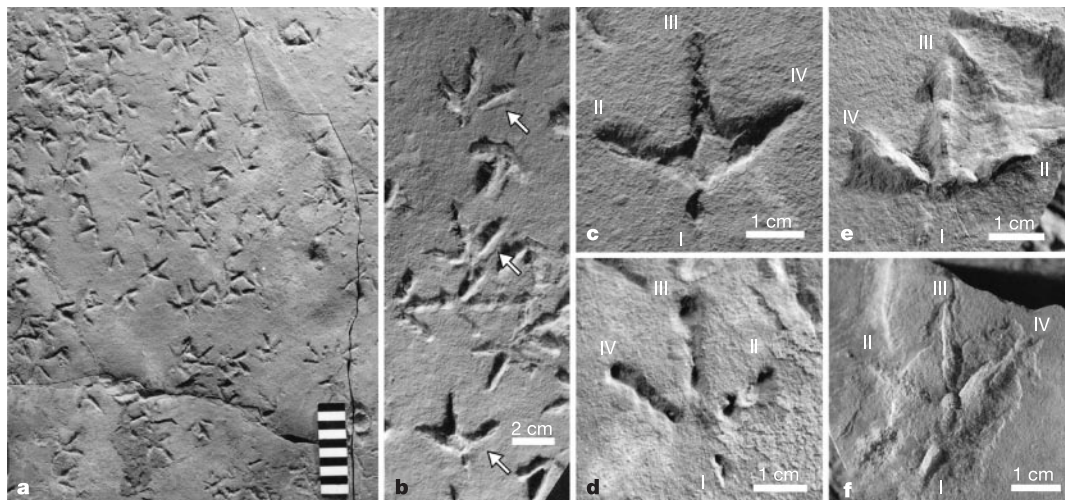


Figure 1 Late Triassic bird-like footprints from the Santo Domingo Formation. **a**, View of part of the lower trampled horizon, showing footprints preserved as natural moulds. Note an increase in footprint density at the upper left of the photo. Scale, 10 cm. **b**, Detailed view of a track preserved as a natural cast and composed of three tetradactyl footprints (arrowed). Note the high pace angulation and rotation of foot axis toward the midline. **c–f** Close-up of four footprint preservational variants, as natural moulds (**c,d**) and casts (**e,f**). All footprints except **e** correspond to the left hind limb. I to IV refer to first to fourth digit imprints. **c**, Deeply

imprinted footprint lacking pad impressions (note remains of matrix within the footprint). **d**, Footprint with well-defined pad impressions and rounded ends of digit prints II to IV. Digit I shows claw mark and a single pad impression. **e**, Footprint with pad impressions, small rounded sole, distal curvature of lateral and medial claw marks away from the foot axis, and shallow hallux impression. Part of a second footprint is shown in the upper right corner. **f**, Probable undertrack that exhibits an elliptical sole and a moderately large hallux impression. Lighting is from the top except for **b**, illuminated from the bottom.

width, 4 mm) and may display a tapered or subparallel outline. The proximal ends of the digit impressions may converge into a conspicuous rounded or elliptical sole (Fig. 1e,f). When present, phalangeal pad impressions are distinct (Fig. 1f) and always display the relation 2-3-4-5 (impressions of toes I-II-III-IV). Toe impressions usually bear slim claw marks, although rounded digit ends have also been recorded. Claw marks in toes II and IV diverge away from the foot axis, whereas that of digit III is parallel to it or curved inward (Fig. 1c,e). Print of digit III is the longest, with those of digits IV, II and I following in decreasing order of length. Divarication of the impression of digits II–IV is high and the angle between the impression of digits II and III is 2°–3° larger, on average, than the angle between the impression of digits III–IV (Fig. 1). The hallux impression is short, thin, usually faint and set at a high angle to digit III (average 160°), usually 2 mm to 8 mm behind the sole or proximal end of digit III; hallux impressions are clearer in the more deeply imprinted tracks than in the shallower tracks. There is no indication of webbing between the digits.

Different criteria have been suggested to distinguish between the footprints of avian and non-avian theropods^{7,8,18–22}. The Santo Domingo tracks described herein meet most proposed features that characterize bird footprints, including (see Fig. 1 and Table 1): (1) an overall similarity to modern bird footprints; (2) footprints that are wider than they are long (not considering the hallux) and of small size; (3) slender digit impressions; (4) a wide angle between digits II and IV; (5) a posterior or posteromedial hallux impression, visible both in shallow and deep tracks; (6) slender claws showing distal curvature of lateral and medial claws away from the foot axis; and (7) a sole or metatarsal–phalangeal impression is visible in some footprints, where digits II to IV converge. Additional indications of an avian affinity are afforded by comparison with tracks of modern waterbirds and waders^{19,21}; these are: (8) high footprint density and absence of preferred orientation; and (9) occurrence in a shallow lacustrine setting, an environment where bird tracks are preferentially preserved. Although these avian features can be found in isolation and, exceptionally, in non-avian footprints, their combined occurrence in the studied track assemblage is exclusive to birds^{2,18,20,22}. This conclusion is supported by the study of a relatively large sample size of footprints preserved in substrates of different consistency, and not merely by single

footprints or short tracks^{2,3}.

Features not in agreement with an avian origin for these bird-like tracks are less significant; they include the presence of distinct pad impressions in some footprints and the absence of associated feeding traces¹⁹. The shallow hallux impression, commonly disconnected with the rest of the foot, suggests that the hallux contacted the ground, but that it was slightly raised and probably not adapted for perching as in some birds (compare with ref. 7). In addition, divarication of digits II–IV in tridactyl non-avian dinosaur footprints may reach high values in a few cases, and high pace angulation is not exclusive to birds^{2,20}.

Bird-like footprints are rare in the Mesozoic era record and occur predominantly in Cretaceous period strata^{19,23}. Pre-Cretaceous evidence is more sparse and can be traced back to the Early Jurassic^{1,3,4,20,24}. For the Triassic period, it has been suggested that some described ichnogenera (mainly *Plesiornis* and *Trisauropodiscus*) might be likewise “aviform” to varying degrees^{1,19,24–26}. However, these footprints show few avian characters and are not comparable with the footprints described herein. *Plesiornis* footprints show low divarication between the impression of digits II and IV (61°–92°), a U-shaped outline, and relatively thick digit imprints⁸, features that contrast with typical bird footprints. The poorly preserved⁹ Triassic specimens of *Trisauropodiscus* are similar to *Anomoepus*^{8,9}; however, that ichnogenus is believed to be the product of an ornithischian trackmaker^{1,8,9}.

Whatever the ichnotaxonomic affinities of these footprints, their producers are unknown from Late Triassic skeletal remains. In particular, the Late Triassic theropod record is sparse²⁷ and no theropod shows evidence of an avian-like reversed hallux⁷. Consequently, these bird-like footprints can only be attributed to an unknown group of theropods showing some avian characters. □

Received 5 February; accepted 8 April 2002; doi:10.1038/nature00818.

1. Haubold, H. *Saurierfahrten* (Ziemsen, Wittenburg, 1984).
2. Thulborn, R. A. *Dinosaur Tracks* (Chapman & Hall, London, 1990).
3. Lockley, M. G. & Hunt, A. P. *Dinosaur Tracks and Other Fossil Footprints of the Western United States* (Columbia Univ. Press, New York, 1995).
4. Lockley, M. G. & Meyer, C. *Dinosaur Tracks and other Fossil Footprints of Europe* (Columbia Univ. Press, New York, 2000).
5. Ostrom, J. H. Bird flight, how did it begin? *Am. Sci.* **67**, 46–56 (1979).
6. Padian, K. & Chiappe, L. M. The origin and early evolution of birds. *Biol. Rev.* **73**, 1–42 (1998).
7. Gatesy, S. M., Middleton, K. M., Jenkins, F. A. & Shubin, N. H. Three-dimensional preservation of foot

- movements in Triassic theropod dinosaurs. *Nature* **339**, 141–144 (1999).
8. Gierliński, G. Avialian theropod tracks from the Early Jurassic strata of Poland. *Zubia* **14**, 79–87 (1996).
 9. Rainforth, E. C. Late Triassic–Early Jurassic dinosaur ichnofaunas, eastern North America and Southern Africa. *J. Vert. Paleol.* **21**, 91A (2001).
 10. Chiappe, L. M. The first 85 million years of avian evolution. *Nature* **378**, 349–355 (1995).
 11. Caminos, R., Zamuner, A. B., Limarino, O. C. & Fauqué, L. El Triásico superior fosilífero en la Precordillera ríojana. *Rev. Asoc. Geol. Arg.* **50**, 262–265 (1995).
 12. Coughlin, T. J. *Linked Orogen–Oblique Fault Zones in the Central Argentine Andes: The Basis of a New Model for Andean Orogenesis and Metallogenesis* PhD thesis (Univ. Queensland, 2001).
 13. Caminos, R. & Fauqué, L. Geologic sheet 2969–II Tinogasta (1:250,000). *Ser. Geol. Min. Arg.* (2001).
 14. Artabe, A. E., Brea, M. & Zamuner, A. B. *Rhexoxylon brunoi* Artabe, Brea et Zamuner, sp. nov., a new Triassic *Corytoposperm* from the Paramillo de Uspallata, Mendoza, Argentina. *Rev. Palaeobot. Palyn.* **105**, 63–74 (1999).
 15. Gradstein, F. M. et al. A Triassic, Jurassic and Cretaceous time scale. *SEPM Spec. Publ.* **54**, 95–126 (1995).
 16. Kent, D. V. & Olsen, P. E. Implications of astronomical climate cycles to the chronology of the Late Triassic. *Zbl. Geol. Paläont.* **8**, 1463–1473 (2000).
 17. Pálfi, J., Smith, P. L. & Mortensen, J. K. A U–Pb and $^{40}\text{Ar}/^{39}\text{Ar}$ time scale for the Jurassic. *Can. J. Earth Sci.* **37**, 923–944 (2000).
 18. Currie, P. Bird footprints from the Gething Formation (Aptian, Lower Cretaceous) of northeastern British Columbia, Canada. *J. Vert. Paleontol.* **1**, 257–264 (1981).
 19. Lockley, M. G., Yang, S.-Y., Matsukawa, M., Fleming, F. & Lim, S.-K. The track record of Mesozoic birds: evidence and implications. *Phil. Trans. R. Soc. Lond. B* **336**, 113–134 (1992).
 20. Fuentes Vidarte, C. Primeras huellas de aves en el Weald de Soria (España). Nuevo icnógeno, *Archaeornithipus* y nueva icnoespecie. *A. mejidi. Est. Geol.* **52**, 63–75 (1996).
 21. Doyle, P., Wood, J. L. & George, G. T. The shorebird ichnofacies: an example from the Miocene of southern Spain. *Geol. Mag.* **137**, 517–536 (2000).
 22. McCrea, R. T. & Sarjeant, W. A. S. in *Mesozoic Vertebrate Life* (eds Tanke, D. H. & Carpenter, K.) 453–478 (Indiana Univ. Press, Bloomington/Indianapolis, 2001).
 23. Lockley, M. G. The vertebrate track record. *Nature* **396**, 429–432 (1998).
 24. Ellenberger, P. Contribution à la classification des pistes de vertébrés du Trias: Les types du Stormberg d'Afrique du Sud (I). *Palaeover. Mém. Extraordinaire*, 1–128 (1972).
 25. Weems, R. E. & Kimmel, P. G. Upper Triassic reptile footprints and a coelacanth fish scale from the Calpeper basin, Virginia. *Proc. Biol. Soc. Wash.* **106**, 390–401 (1993).
 26. Gierliński, G. & Sabath, K. Protoavian affinity of the *Plesiomis* trackmaker. *J. Vert. Paleol.* **18**, 46A (1998).
 27. Heckert, A. B. & Lucas, S. G. Global correlation of the Triassic theropod record. *Gaia* **15**, 63–74 (1998).

Acknowledgements

This work was funded by the Agencia Nacional de Promoción Científica y Tecnológica of Argentina. M. Archangelsky helped during fieldwork. We thank T. Coughlin for sharing unpublished geochronological data. W.A.S. Sarjeant critically read the manuscript and A. Zamuner commented on biostratigraphical implications of the fossil flora.

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to R.N.M. (e-mail: melchor_r@cpsarg.com).

Overlap of internal models in motor cortex for mechanical loads during reaching

Paul L. Gribble* & Stephen H. Scott

Department of Anatomy and Cell Biology, CIHR Group in Sensory-Motor Systems, Centre for Neuroscience Studies, Queen's University, Kingston, Ontario K7L 3N6, Canada

A hallmark of the human motor system is its ability to adapt motor patterns for different environmental conditions, such as when a skilled ice-hockey player accurately shoots a puck with or without protective equipment. Each object (stick, shoulder pad, elbow pad) imparts a distinct load upon the limb, and a key problem in motor neuroscience is to understand how the brain controls movement for different mechanical contexts^{1,2}. We

addressed this issue by training non-human primates to make reaching movements with and without viscous loads applied to the shoulder and/or elbow joints, and then examined neural representations in primary motor cortex (MI) for each load condition. Even though the shoulder and elbow loads are mechanically independent, we found that some neurons responded to both of these single-joint loads. Furthermore, changes in activity of individual neurons during multi-joint loads could be predicted from their response to subordinate single-joint loads. These findings suggest that neural representations of different mechanical contexts in MI are organized in a highly structured manner that may provide a neural basis for how complex motor behaviour is learned from simpler motor tasks.

Behavioural studies suggest that the brain uses internal models—neural processes that mimic the characteristics of the body or environment—to predict and generate motor commands for movement^{1,3}, but little is known about neural computations associated with these representations^{4,5}. Here we test two qualitatively distinct hypotheses about the organization in the brain of internal models for different mechanical loads². One possibility is that internal models for different loads are represented within a single controller that encapsulates all possible loads (Fig. 1a). A second possibility is a more modular scheme in which multiple controllers co-exist, each suitable for one context (or a small set of contexts) (Fig. 1b). These two hypotheses predict differences in how individual neurons in the brain respond when loads are applied in a given motor task: either a cell consistently changes activity for all mechanical loads (former), or it changes activity only for one or a subset of loads (latter). We tested these two alternatives using a reaching task with different dynamic loads, and recorded neuronal activity in MI, a region intimately involved in volitional motor control where cells often respond to changes in force output^{4–8}.

We trained monkeys to wear a robotic exoskeleton (KINARM) that permitted horizontal limb movements using flexion and extension motions at the shoulder and elbow^{9,10}. Monkeys made reaching movements without loads (NL) and with one of three

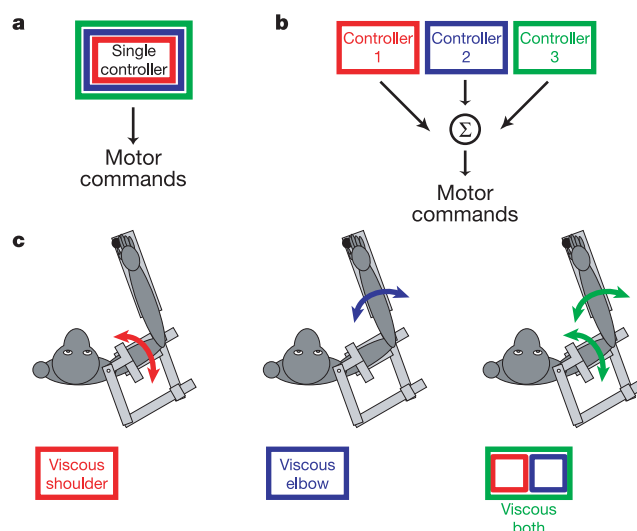


Figure 1 Two hypotheses about the neural control of different mechanical loads. A single-controller that encapsulates all load contexts (a), or multiple controllers, each of which represent individual loads (b). c, Experimental design used to assess the neural representation of multiple loads. A robotic exoskeleton applied velocity-dependent (viscous) loads to the monkey's arm during reaching movements. Viscous shoulder (VS) and viscous elbow (VE) are mechanically independent loads. Viscous both (VB) is the superposition of viscous shoulder and viscous elbow loads.

* Present address: Department of Psychology, The University of Western Ontario, London, Ontario N6A 5C2, Canada

velocity-dependent loads (Fig. 1c): one where loads are applied at the shoulder only (viscous shoulder, VS), one where loads are applied at the elbow only (viscous elbow, VE), and a superposition of the two single-joint loads (viscous both, VB). Single-joint loads were created easily with this device, as forces are applied directly on the limb segments rather than at the hand. VS and VE loads allow us to address whether mechanically independent loads are represented by distinct populations of neurons or distributed across a single population in MI. Moreover, because VE and VS are both mechanically subordinate to VB, we can also explore whether or not the representation of a complex load (VB) can be generated from simpler representations (VE and VS). A key to these studies was to ensure that the kinematics of reaching remained similar between mechanical conditions, so that changes in neural activity could be related to a representation of the mechanical load and not features of movement kinematics (Supplementary Information).

A total of 101 cells in primary motor cortex were recorded from three monkeys (see Methods) in the NL condition and all three load conditions (monkey A, 44; monkey B, 18; monkey C, 39). Over half of these cells (55 of 101) showed significant changes in firing rate relative to their activity in NL, for at least one movement direction and one load condition ($P < 0.05$, analysis of variance, ANOVA; monkey A, 28; monkey B, 13; monkey C, 14). These changes in activity are not solely due to afferent feedback because 19 cells showed significant changes in firing rate before movement onset (150–0 ms before movement onset). Figure 2 shows typical

responses for two cells, one that increased its activity (a, c, e) and another that decreased its activity (b, d, f) for VE and VB.

Although the presence of the viscous loads required an increase in muscle force to oppose the load (see below), half of the cells that showed statistically significant changes in mean firing rate decreased their mean discharge during movement (Fig. 3a). These decreases in discharge may partially reflect reductions in co-contraction of antagonistic muscles with loads^{11,12}. As observed in previous load studies⁴, changes in activity with load (increases or decreases) tended to occur along each cell's preferred direction, the direction of movement with maximal activity during unloaded reaching (Fig. 3b).

We found that 48 cells responded to at least one of the two single-joint loads, with 8, 26 and 14 cells responding to VS only, VE only or to both loads, respectively. The proportion of cells responding to both single-joint loads is larger than predicted if each load was randomly represented across our cell sample ($\chi^2 = 6.8$, $P < 0.01$). Moreover, there was a highly consistent relationship between how a cell changed its activity for VS and VE ($n = 48$, $r^2 = 0.65$, $P < 0.001$). Cells that increased discharge for VS also tended to increase discharge for VE, while decreases were likewise associated (top projection plane in Fig. 4a).

An important point is that 31 of the 38 cells that responded to the multi-joint load (VB) also responded to either of the mechanically subordinate single-joint loads (VS or VE). This is much higher than expected if single and multi-joint loads were each randomly

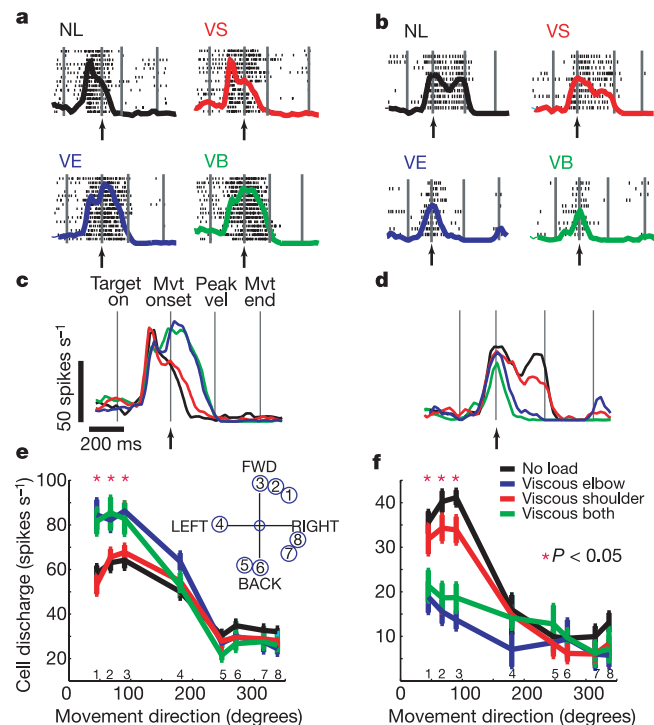


Figure 2 Neural activity of two cells during movement in different dynamic loads. **a, b**, Each set of rasters shows the discharge of a cell in each of 15 (**a**) or 10 (**b**) repeat trials in the no-load condition and in each viscous load, for a single reaching movement forward (target 3, see **e** inset). Vertical bars denote (from left to right) target onset, movement onset, peak velocity and movement end. Data are aligned to movement onset, indicated by vertical arrows. **c, d**, Mean discharge over time in the four load conditions, computed using a 50-ms sliding window every 20 ms. **e, f**, Mean discharge between 150 ms before movement onset and peak velocity in each load condition, for movements to eight targets (see inset, **e**). Vertical bars indicate one standard error. Red asterisk, $P < 0.05$. **a, c, e**, Shows a cell (no. 034) from monkey B that increased discharge for VE and VB loads; **b, d, f** shows a cell (no. 017) from monkey A that decreased discharge for VE and VB loads.

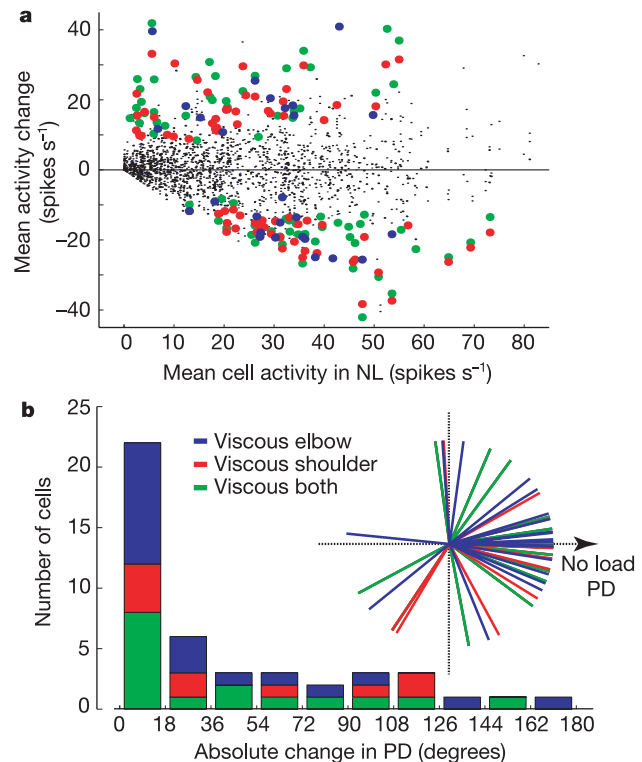


Figure 3 Changes in cell activity across three load conditions. **a**, Mean discharge between 150 ms before movement onset to peak velocity during unloaded reaching movements (NL) versus change in activity between unloaded and loaded conditions. Large red, blue and green dots denote significant changes in activity associated with VE, VS and VB conditions, respectively ($P < 0.05$, ANOVA). Small black dots denote non-significant changes. **b**, Histogram of the distribution of load-related preferred directions (PD) relative to each cell's PD for unloaded reaching. Inset panel illustrates this distribution using individual lines of orientation for each load-related PD relative to each cell's PD for unloaded reaching, which has been arbitrarily rotated to the right in the figure.

represented across our cell sample ($\chi^2 = 28.3$, $P < 0.01$). Moreover, a categorical analysis such as ANOVA fails to capture fully the consistency of the coupling between each cell's response for VB and its associated responses for VS and VE. Figure 4 illustrates this robust linkage; the observed change in discharge (relative to NL) for VB is plotted against the corresponding changes in discharge for VS and VE. A notable result is that almost all cells (45/55, 82%) are located in only two of the eight possible quadrants ($\chi^2 = 98.0$, $P < 0.001$). These two quadrants reflect cell responses where cells either increased (22 cells, green dots) or decreased (23 cells, red dots) their discharge compared to NL in all three load conditions.

We have previously found that multi-joint postural loads could be predicted from single-joint loads using a vector summation model which supposes that neural activity can be represented by a

coordinate frame with the shoulder and elbow each representing orthogonal axes¹³. This vector model also predicted the response of neurons during VB based on each cell's response during VS and VE conditions ($r^2 = 0.81$, $P < 0.001$; slope of best fit similar to unity, $P > 0.05$; Fig. 4b). A linear summation model was also attempted, in which the change in discharge for VB was assumed to equal the sum of the changes in VS and VE. Although there was a consistent and statistically significant relationship between predicted and actual cell discharge using the linear model ($r^2 = 0.81$, $P < 0.001$), this model tended to overestimate cell discharge in VB (paired t -test, $P < 0.01$; slope of best fit 1.6, $P < 0.0001$).

Electromyographic (EMG) activity of forelimb muscles shared some but not all features observed for MI cells (Supplementary Information). Unlike MI cells, almost all muscles (55/57) changed activity for at least one mechanical load, and 24 changed activity for all three. In addition, EMG changes were predominantly (78%) increases in activity, whereas there was almost an equal number of neurons that either increased or decreased their activity with loads ($\chi^2 = 18.2$, $P < 0.001$). Thus our observations in MI cannot be explained solely on the basis of muscle activation patterns.

The present results provide details on how internal models for different velocity-dependent loads are represented in the discharge patterns of cells in MI. In theory, the multi-joint and the two single-joint loads could each have been represented by separate population of neurons (that is, separate neural modules). However, we found substantive overlap in representations, particularly between multi-joint and subordinate single-joint loads, suggesting that neural activity in MI behaves like a single controller (that is, a single internal model) for these different contexts. Further, there was a partial overlap in the representations of the two single-joint load conditions that may partially reflect the coarse somato-motor map in primary motor cortex¹⁴ and the fact that some motor-cortical cells are related to multiple muscles spanning different joints¹⁵. Although other regions of the brain, such as cerebellum, may use a more modular organization for multiple mechanical contexts^{16,17}, the output from these other regions is presumably integrated to form the representations that we observed in MI. Finally, our results suggest that internal models for simpler loads are combined to form models for more complex loads, and may explain why training on simpler tasks can be transferred and improve performance on related, more complex tasks^{18,19}. □

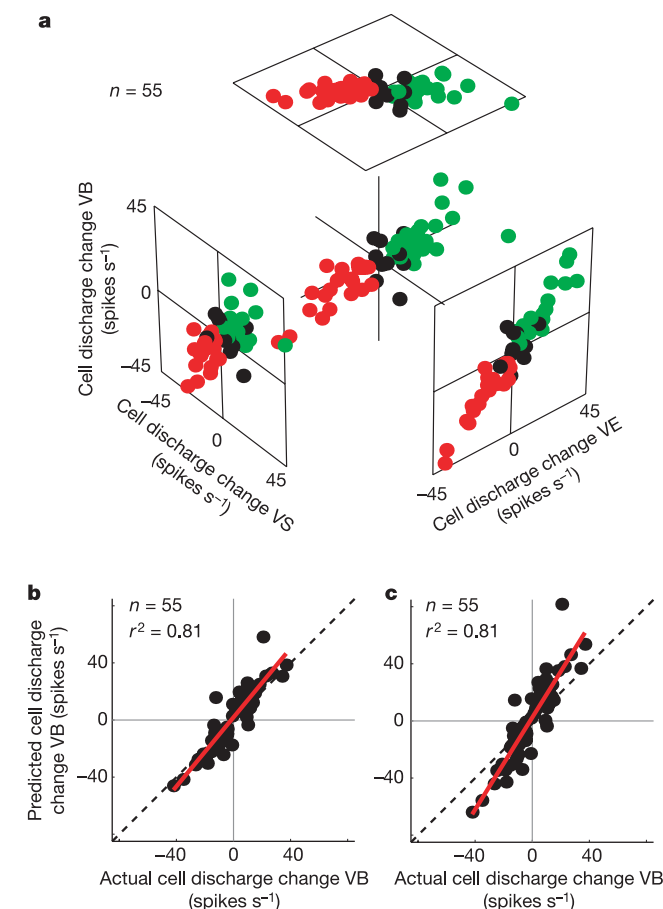


Figure 4 Neural representation of mechanically dependent loads. **a**, Relationship between cell discharge change (relative to no load) in VS, VE and VB loads for the 55 cells (out of 101 recorded in 3 monkeys in all 4 conditions) that showed significant changes in discharge in at least one movement direction for at least one load condition (VS, VE or VB). Neural responses were highly consistent across the different load conditions. Significant relationships were observed between cell discharge change in VS versus VE ($r^2 = 0.66$, $P < 0.0001$), VS versus VB ($r^2 = 0.62$, $P < 0.0001$) and VE versus VB ($r^2 = 0.83$, $P < 0.0001$). Data plotted for each cell represent the movement direction showing the greatest absolute change in mean discharge. Green dots, data from cells that increased discharge in all three loads; red dots, cells that decreased discharge in all three loads; black dots, cells that showed other patterns. **b**, Predicting neural activity for a multi-joint load. The change in discharge (from NL) in VB is plotted against predicted discharge change using a vector combination of neural activity in individual single-joint component loads ($\Delta VB = (\Delta VS^2 + \Delta VE^2)^{1/2}$). The model successfully predicts a neuron's response in VB using a combination of its activity in VS and VE ($r^2 = 0.81$, $P < 0.001$). Solid red line is line of best fit, dashed line represents $y = x$. A linear model (**c**) tended to overestimate cell discharge in VB.

Methods

Experiments

Three male rhesus monkeys (*Macaca mulatta*, 6–7 kg) were trained to wear a robotic exoskeleton (KINARM) and move from a central target to one of eight peripheral targets located on the circumference of a circle^{9,10}. Training included experimental conditions NL, VS, VE and VB (see text). In all cases, the torques applied by KINARM were negatively proportional to joint velocity, and opposed joint rotation, and no load was applied when the limb was stationary. Monkeys were fully trained to perform these movements consistently in all load conditions before neural recordings were initiated. Rigorous spatial and temporal accuracy requirements were imposed during both training and recording sessions to ensure that movement kinematics were held constant across all load conditions (Supplementary Information). All experimental procedures were approved by the Queen's University Animal Care Committee.

Data recording

Recording chambers were surgically implanted under isoflurane anaesthesia, and conventional techniques were used for extracellular recording of single-neuron activity in the proximal arm representation of primary motor cortex contralateral to the arm used to perform reaching movements^{10,20}. Cell activity and several measures of limb motion were typically recorded in six repeat trials for each of the movement targets in a pseudo-random block design. In subsequent data analyses, the first movement to each target was excluded. In the first two monkeys (A and B), neural activity was recorded for movement to eight targets, selected specifically to sample roughly equal intervals in joint-torque space. These directions in torque-space correspond to the following directions in hand-space: 45°, 67.5°, 90°, 180°, 247.5°, 270°, 305° and 327.5°, where 0° is to the right, and positive rotation is anticlockwise (see Fig. 2e and Supplementary Information for diagram). Initially a block of 48 trials (8 targets × 6 repeat trials) was recorded in the NL condition, followed by three blocks of 48 trials in the VS, VE and VB conditions, respectively, where the order of the three loads was randomized for each cell. Finally, as a control for time-dependent changes in cell activity, a second block of trials in the NL condition was recorded after the

completion of the three viscous load conditions (Supplementary Information). Monkey C was trained to perform movements with loads to eight directions uniformly distributed in hand-space. In order to save time during the viscous load conditions, neural activity was recorded only to targets at 90° and 270°, directions associated with large shoulder and elbow rotations, and thus large loads (Supplementary Information).

Data analysis

Cell activity was represented as the number of action potentials recorded between 150 ms before movement onset (defined on the basis of 5% of peak tangential hand velocity) and peak tangential hand velocity. Muscle torques at the shoulder and elbow were computed using inverse dynamics⁹. Differences in mean cell discharge were assessed between load conditions and across movement directions using two-way ANOVA followed by individual comparisons between means using a Bonferroni correction for multiple comparisons. For monkeys A and B, differences were considered significant at $P < 0.0021$ ($P < 0.05 / 24$ comparisons: 8 target directions $\times$ 3 load conditions); for monkey C differences were considered significant at $P < 0.0083$ ($P < 0.05 / 6$ comparisons: 2 targets $\times$ 3 loads). All probability values reported in the text represent Bonferroni-corrected values.

The preferred direction (PD), the direction of maximal activity, of each cell for unloaded reaching movements was computed^{10,21}. We also defined a load-related PD based on the absolute change in cell activity between NL and each load condition. In this case, the preferred direction signified the largest change (either increase or decrease) in activity. Bootstrap techniques were used to identify whether these directional signals were statistically significant ($P < 0.01$)^{10,20}.

In each monkey, the EMG activity of up to 16 forelimb muscles (mono- and bi-articular muscles spanning shoulder and elbow) was recorded during conditions NL, VB, VS and VE. Pairs of single-strand wires were inserted percutaneously in monkeys A, B and C, and pairs of multi-strand wires were implanted chronically in monkeys A and C under aseptic conditions^{10,22}. Some muscles were sampled twice in monkeys A and C, once using chronically implanted wires and a second time using percutaneously inserted wires providing a total of 57 samples of EMG (Supplementary Information).

Received 25 January; accepted 16 April 2002; doi:10.1038/nature00834.

1. Wolpert, D. M. & Ghahramani, Z. Computational principles of movement neuroscience. *Nature Neurosci.* **3** suppl., 1212–1217 (2000).
2. Wolpert, D. M. & Kawato, M. Multiple paired forward and inverse models for motor control. *Neural Netw.* **11**, 1317–1329 (1998).
3. Shadmehr, R. & Mussa-Ivaldi, F. A. Adaptive representation of dynamics during learning of a motor task. *J. Neurosci.* **14**, 3208–3224 (1994).
4. Kalaska, J. F., Cohen, D. A., Hyde, M. L. & Prud'homme, M. A comparison of movement direction-related versus load direction-related activity in primate motor cortex, using a two-dimensional reaching task. *J. Neurosci.* **9**, 2080–2102 (1989).
5. Gandolfo, F., Li, C., Benda, B. J., Schioppa, C. P. & Bizzi, E. Cortical correlates of learning in monkeys adapting to a new dynamical environment. *Proc. Natl Acad. Sci. USA* **97**, 2259–2263 (2000).
6. Kalaska, J. F., Scott, S. H., Cisek, P. & Sergio, L. E. Cortical control of reaching movements. *Curr. Opin. Neurobiol.* **7**, 849–859 (1997).
7. Porter, R. & Lemon, R. *Corticospinal Function and Voluntary Movement* (Oxford Univ. Press, Oxford, 1995).
8. Ashe, J. Force and the motor cortex. *Behav. Brain Res.* **87**, 255–269 (1997).
9. Scott, S. H. Apparatus for measuring and perturbing shoulder and elbow joint positions and torques during reaching. *J. Neurosci. Methods* **89**, 119–127 (1999).
10. Scott, S. H., Gribble, P. L., Graham, K. M. & Cabel, D. W. Dissociation between hand motion and population vectors from neural activity in motor cortex. *Nature* **413**, 161–165 (2001).
11. Humphrey, D. R. & Reed, D. J. In *Motor Control Mechanisms in Health and Disease* Advances in Neurology no. 39 (ed. Desmedt, J.) 347–372 (Raven, New York, 1983).
12. Fetz, E. E., Cheney, P. D., Mewes, K. & Palmer, S. In *Peripheral Control of Posture and Locomotion* (eds Allum, J. A. H. & Hulliger, M.) 437–449 (Elsevier, New York, 1989).
13. Cabel, D. W., Cisek, P. & Scott, S. H. Neural activity in primary motor cortex related to mechanical loads applied to the shoulder and elbow during a postural task. *J. Neurophysiol.* **86**, 2102–2108 (2001).
14. Sanes, J. N. & Schieber, M. H. Orderly somatotopy in primary motor cortex: does it exist? *Neuroimage* **13**, 968–974 (2001).
15. McKiernan, B. J., Marcario, J. K., Karrer, J. H. & Cheney, P. D. Corticomotorneural postspike effects in shoulder, elbow, wrist, digit, and intrinsic hand muscles during a reach and prehension task. *J. Neurophysiol.* **80**, 1961–1980 (1998).
16. Imamura, H. et al. Human cerebellar activity reflecting an acquired internal model of a new tool. *Nature* **403**, 192–195 (2000).
17. Wolpert, D. M., Miall, R. C. & Kawato, M. Internal models in the cerebellum. *Trends Cogn. Sci.* **2**, 338–347 (1998).
18. Schmidt, R. A. & Wrisbert, C. A. *Motor Learning and Performance: A Problem-Based Learning Approach* (Human Kinetics, Champaign, 2000).
19. Wightman, D. C. & Lintern, G. Part-task training for tracking and manual control. *Hum. Factors* **27**, 267–283 (1985).
20. Scott, S. H. & Kalaska, J. F. Reaching movements with similar hand paths but different arm orientations. I. Activity of individual cells in motor cortex. *J. Neurophysiol.* **77**, 826–852 (1997).
21. Gribble, P. L. & Scott, S. H. Method for assessing directional characteristics of non-uniformly sampled neural activity. *J. Neurosci. Methods* **113**, 187–197 (2002).
22. Loeb, G. E. & Gans, C. *Electromyography for Experimentalists* (Univ. Chicago Press, Chicago, 1986).

Supplementary Information accompanies the paper on Nature's website (<http://www.nature.com/nature>).

Acknowledgements

We thank K. Moore for technical assistance, and D.W. Cabel and S. Chan who assisted in some of the training and neuronal recording sessions. We thank D. Munoz, M. Pare and K. Rose for comments on this manuscript. This work was supported by a CIHR grant and scholarship (S.H.S.) and a CIHR postdoctoral fellowship (P.L.G.).

Competing interests statement

The authors declare competing financial interests: details accompany the paper on Nature's website (<http://www.nature.com>).

Correspondence and requests for materials should be addressed to S.H.S. (e-mail: steve@biomed.queensu.ca).

Oligodendrocyte-myelin glycoprotein is a Nogo receptor ligand that inhibits neurite outgrowth

Kevin C. Wang^{*†}, Vuk Koprivica^{*†}, Jieun A. Kim^{*}, Rajeev Sivasankaran^{*}, Yong Guo[‡], Rachel L. Neve^{‡§} & Zhigang He^{*†}

^{*} Division of Neuroscience, Children's Hospital, and [†] Program in Neuroscience, Harvard Medical School, 320 Longwood Avenue, Boston, Massachusetts 02115, USA

[‡] Aventis Pharmaceuticals, Bridgewater, New Jersey 08807, USA

[§] McLean Hospital, 115 Mill Street, Belmont, Massachusetts 02478, USA

The inhibitory activity associated with myelin is a major obstacle for successful axon regeneration in the adult mammalian central nervous system (CNS)^{1,2}. In addition to myelin-associated glycoprotein (MAG)^{3,4} and Nogo-A^{5–7}, available evidence suggests the existence of additional inhibitors in CNS myelin⁸. We show here that a glycosylphosphatidylinositol (GPI)-anchored CNS myelin protein, oligodendrocyte-myelin glycoprotein (OMgp), is a potent inhibitor of neurite outgrowth in cultured neurons. Like Nogo-A, OMgp contributes significantly to the inhibitory activity associated with CNS myelin. To further elucidate the mechanisms that mediate this inhibitory activity of OMgp, we screened an expression library and identified the Nogo receptor (NgR)⁹ as a high-affinity OMgp-binding protein. Cleavage of NgR and other GPI-linked proteins from the cell surface renders axons of dorsal root ganglia insensitive to OMgp. Introduction of exogenous NgR confers OMgp responsiveness to otherwise insensitive neurons. Thus, OMgp is an important inhibitor of neurite outgrowth that acts through NgR and its associated receptor complex. Interfering with the OMgp/NgR pathway may allow lesioned axons to regenerate after injury *in vivo*.

To examine whether any GPI-linked proteins in CNS myelin act as inhibitors of neurite outgrowth, we treated purified myelin from bovine white matter with phosphatidylinositol-specific phospholipase C (PI-PLC) and examined the released proteins for their ability to alter growth cone morphology in an assay of growth cone collapse using dorsal root ganglia from chicks at embryonic day 13 (E13 DRG)^{6,9,10}. PI-PLC-released proteins from CNS myelin exhibited marked growth cone collapsing activity (Fig. 1). By SDS-polyacrylamide gel electrophoresis (PAGE) and silver staining, we found that a band of relative molecular mass about 110,000 ($M_r \sim 110K$) was significantly enriched in this fraction (Fig. 2a). Because the size of this band was similar to that of a previously identified CNS myelin protein, OMgp^{11,12}, we used anti-OMgp antibodies to detect enrichment of cleaved OMgp in the PI-PLC supernatants by western blot. Anti-OMgp antibodies detected a band of comparable size in the PI-PLC-treated supernatants

(Fig. 2b), suggesting that OMgp is a component released from CNS myelin by PI-PLC. Next, we examined whether purified recombinant OMgp protein was able to inhibit neurite outgrowth in both the growth cone collapse^{6,9,13} and the neurite outgrowth assays^{3-7,9}. Like the PI-PLC-treated myelin supernatants, purified recombinant polyhistidine-tagged OMgp protein (OM-His), but not proteins purified from COS-7 cells transfected with a control vector (data not shown), induced the collapse of E13 DRG growth cones dose dependently. The effective concentration for half-maximal response (EC_{50}) for OM-His was about 1.5 nM (Fig. 2c). Consistent with the notion that it is an inhibitor of neurite outgrowth, the OMgp protein is highly expressed by mature oligodendrocytes positive for myelin basic protein (MBP)¹⁴, and enriched in the axon-adjacent myelin layers^{11,12}. Moreover, OM-His, when presented either as an immobilized substrate or in a soluble form (Fig. 2d, e), inhibited neurite outgrowth of cerebellar granule neurons (CGNs) from rats at postnatal day 7-9 (P7-9), also dose dependently. The OMgp-induced inhibitory responses were similar to that brought about by treatment of the same neurons with an alkaline phosphatase fusion protein containing the 66-amino-acid extracellular domain of Nogo-A (AP-66)^{6,9} (Fig. 2c-e). This suggests that OMgp, like Nogo-66, is a potent neurite outgrowth inhibitor.

To further examine the functional importance of OMgp as a CNS myelin-associated inhibitor, we used peanut agglutinin (PNA)-agarose beads¹¹ to specifically deplete OMgp from solubilized CNS myelin. Although the inhibitory activity in the PNA-depleted myelin was significantly lower than that in control myelin, the OMgp-enriched eluates from the same PNA column displayed a potent inhibitory activity (see Supplementary Information). By using a combination of PNA-agarose¹¹ and Q-Sepharose^{3,15,16} columns, we generated myelin fractions enriched in MAG, Nogo-A or OMgp. Comparisons of the EC_{50} for each of these fractions show that the OMgp-enriched fraction inhibited neurite outgrowth to a similar extent as the Nogo-A-enriched fraction, but more than the MAG-enriched fraction (see Supplementary Information).

To investigate the mechanisms by which OMgp inhibits neurite outgrowth, we used an AP-OMgp fusion protein (AP-OM) as a probe to identify OMgp-binding proteins on the cell surface by an expression cloning strategy^{9,13,17}. We examined pools of complementary DNA from an adult human brain cDNA expression

library, and isolated two cDNAs that encoded OMgp-binding proteins. Sequence analysis revealed that both cDNAs contained the full-length coding region of the Nogo-66 receptor (NgR), a high-affinity receptor for the extracellular domain of Nogo-A (ref. 9). We then determined the binding affinity of expressed NgR for AP-OM as 5 nM, similar to that determined for Nogo-66 (7 nM)⁹. These data suggest that NgR is a protein that binds OMgp with high affinity. We next performed co-precipitation experiments by incubating glutathione S-transferase (GST) or a GST fusion protein containing the entire extracellular domain of NgR (GST-NgR) with or without OM-His. GST-NgR (Fig. 3a, lane 6), but not

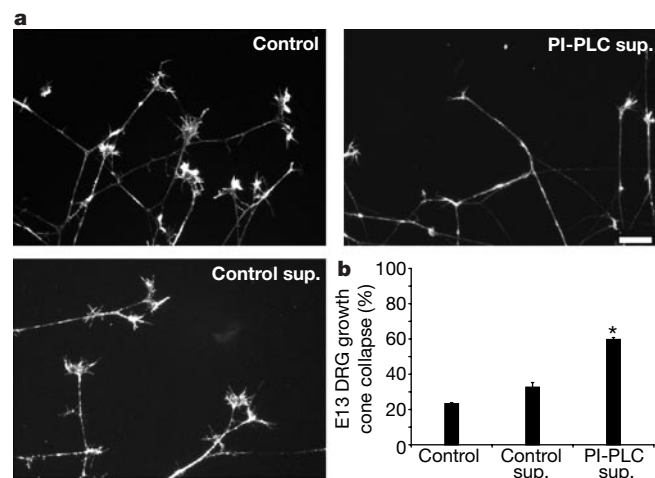


Figure 1 GPI-anchored myelin component(s) induces growth cone collapse. **a**, Chick E13 DRG explants exposed to nothing (control), and the supernatant of CNS myelin-treated with (PI-PLC sup.) or without (control sup.) PI-PLC. Scale bar, 40 μ m. **b**, Quantification of the results shown in **a**. Asterisk, collapse significantly different ($P < 0.005$, Student's t -test) from control.

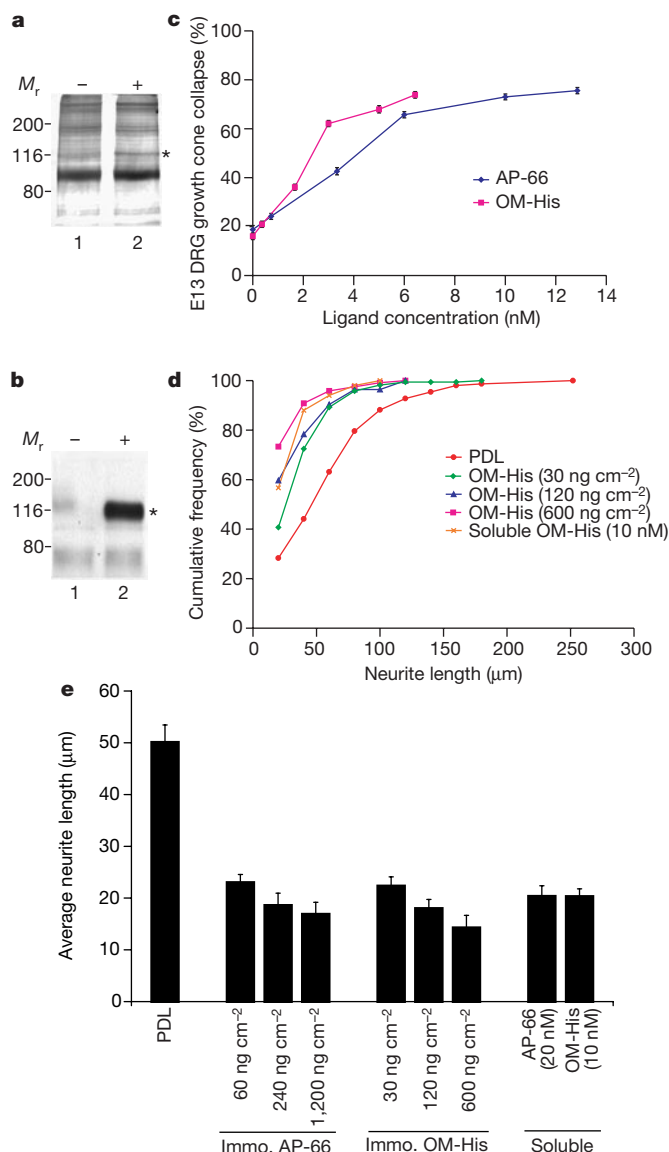


Figure 2 OMgp is a myelin-associated inhibitor. **a**, Silver staining of untreated (–, lane 1) and PI-PLC-treated (+, lane 2) protein supernatants. **b**, Western blot of the same samples probed with anti-OMgp antibodies. **c**, Quantification of the results of E13 chick DRG growth cone collapse assays. Estimated EC_{50} : 2.3 nM (AP-66) and 1.5 nM (OM-His). **d**, Distributions of rat CGNs with different neurite lengths plotted as cumulative histograms. **e**, Neurite lengths (mean $\pm$ s.e.m.) quantified from neurons cultured on immobilized substrates or in the presence of soluble recombinant proteins. Statistical analysis was done by one-way analysis of variance (ANOVA) ($F = 48.78$, d.f. = 26, $P < 0.001$) using a post-hoc Fisher protected test. All groups are significantly different from the control PDL group in **d** and **e**.

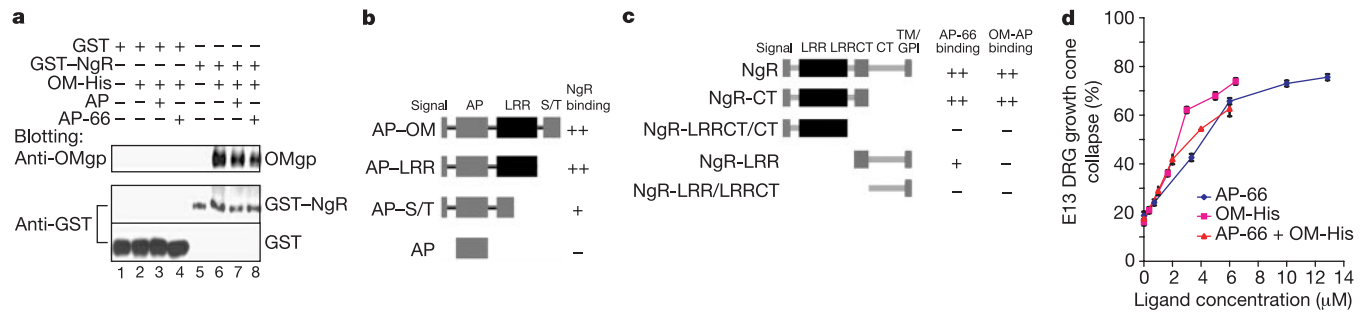


Figure 3 The Nogo-66 receptor is an OMgp-binding protein. **a**, Co-precipitation of GST or GST-NgR with OM-His in the presence of AP or AP-66. **b**, Summary diagram of different AP-OMgp fusion proteins binding to NgR-expressing cells. Signal, signal peptide. **c**, Summary diagram of AP-66 and OM-AP binding to COS-7 cells expressing different

truncations of NgR. CT, unique C-terminal region; TM/GPI, transmembrane domain/GPI anchor. **d**, Growth cone collapse responses of E13 chick DRG induced by AP-66 alone, OM-His alone, or AP-66 plus OM-His (estimated EC_{50} are 2.3 ± 1.3 nM, 1.6 ± 1.3 nM and 2.5 ± 1.1 nM for these treatments, respectively).

the control GST protein (lane 2), bound to OM-His, suggesting a direct interaction between NgR and OMgp. We further determined which domain(s) of OMgp was responsible for binding to NgR. Like NgR, OMgp is a GPI-linked protein containing a leucine-rich repeat (LRR) domain. An AP fusion protein containing only the LRR domain of OMgp (AP-LRR) was found to be sufficient to bind strongly to NgR-expressing cells (Fig. 3b). In addition, the carboxy-terminal domain with serine-threonine repeats (AP-S/T) alone was also able to bind weakly to NgR-expressing cells (Fig. 3b).

To explore how NgR interacts at the molecular level with these two inhibitors, we made a series of deletion constructs of NgR. We found that, although both the LRR and C-terminal LRR (LRRCT) domains of NgR were required for binding to OMgp, the LRRCT domain alone was sufficient for binding to Nogo-66 (Fig. 3c). Thus, OMgp and Nogo-66 bind to overlapping regions of NgR. In assays of cell-surface binding (data not shown) and co-precipitation (Fig. 3a), AP-66 and OM-His proteins consistently competed for binding to NgR. To examine the functional consequences of the molecular interaction of the two ligands with NgR, we compared the collapsing activity of OM-His plus AP-66 with that of OM-His or AP-66 alone (Fig. 3d). The estimated EC_{50} for OM-His

plus AP-66 (2.5 nM) was similar to that for OM-His (1.5 nM) and AP-66 (2.3 nM), suggesting a redundant effect between OM-His and AP-66 in inducing growth cone collapse. Because the binding affinities of both OMgp and Nogo-66 to NgR are similar, our results imply that these two myelin components act independently through NgR to inhibit neurite outgrowth.

Because the GPI-linked NgR protein can be released by PI-PLC, we next examined whether PI-PLC treatment could affect axonal responsiveness to OMgp. Consistent with a previous study⁹, PI-PLC treatment did not alter the growth cone morphology of E13 chick DRG neurons, but rendered these axons insensitive to Nogo-66 (Fig. 4a). Similarly, PI-PLC treatment abolished the growth cone collapsing activity of OMgp (Fig. 4a). As a control, the growth cone collapsing activity of Semaphorin 3A^{10,13}, known to be mediated by transmembrane receptor molecules including neuropilin-1 and members of the plexin family¹⁸, was not affected by PI-PLC treatment (Fig. 4a). Even though PI-PLC also cleaves other GPI-anchored proteins on the axonal surface, these results support the notion that GPI-anchored proteins, most probably NgR, act as necessary signal transducers of the inhibitory activity of OMgp.

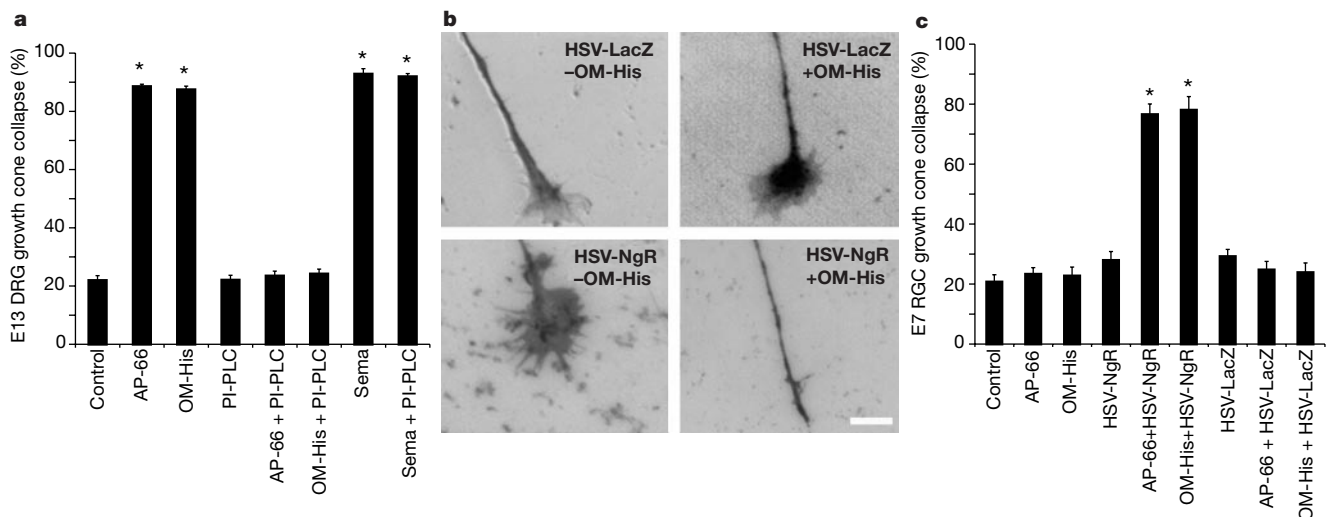


Figure 4 NgR mediates OMgp-elicited growth cone collapse. **a**, Results of growth cone collapse assays on E13 chick DRG. Statistical analysis was by one-way ANOVA ($F = 1,249$, d.f. = 23, $P < 0.0001$), using a post-hoc Fisher protected test. **b**, Collapse responses to OM-His of chick E7 RGC explants infected with HSV-LacZ or HSV-NgR. Scale

bar, 5 μm. **c**, Quantified results from **b**. Statistical analysis was by one-way ANOVA ($F = 126.6$, d.f. = 26, $P < 0.0001$), using a post-hoc Fisher protected test. Asterisks in **a** and **c** mark collapse percentages significantly different from untreated controls.

To assess whether NgR is capable of mediating OMgp-induced inhibitory activity on neurite outgrowth, we took a gain-of-function approach to examine whether expression of NgR conferred OMgp responsiveness to otherwise insensitive neurons. Chick E7 retinal ganglion neurons (RGN) are insensitive to Nogo-66, but introduction of exogenous NgR in these neurons renders their growth cones responsive to Nogo-66 (ref. 9). Using the same strategy, we made a recombinant herpes simplex virus (HSV) that drives expression of a Flag-tagged full-length human NgR (Flag-NgR) in infected neurons. Upon infection, 80% of the E7 RGNs expressed the Flag-NgR protein, as assessed by immunocytochemistry with an anti-Flag antibody. No significant morphological changes were observed in the HSV-infected neurons (Fig. 4b, c). Consistent with a previous study⁹, expression of Flag-NgR, but not control β -galactosidase, conferred a growth cone collapse response to Nogo-66 in E7 RGN (Fig. 4c). Furthermore, the growth cones of NgR-expressing axons also became collapsible by OMgp (Fig. 4b, c). Taken together, our results suggest that, like Nogo-66, OMgp acts through NgR and its associated receptor complex to inhibit axon outgrowth. In contrast to Nogo-A, most of which is localized intracellularly^{5–7}, OMgp is predominantly localized on the surfaces of oligodendrocytes and axon-adjacent myelin layers^{11,12,14}, suggesting that OMgp is a physiological ligand of NgR. Although the precise contributions of OMgp and Nogo-A in restricting axon regeneration *in vivo* remain to be determined, our results suggest that multiple inhibitors may exert their effects through the same signalling pathway, which may provide crucial insights into stimulating axon regeneration after human CNS injury. □

Methods

Purification, PI-PLC treatment and OMgp depletion of myelin

Myelin was prepared from white matter of bovine brain according to established protocols¹⁹. For PI-PLC treatment, aliquots of myelin suspensions in water (10 mg ml^{-1}) were incubated with or without 2.5 U ml^{-1} PI-PLC (Sigma) at 37°C for 2 h, before centrifugation (360,000g for 60 min). The supernatants were concentrated, partitioned in Triton X-114, and used for assays and for detection by western analysis.

Expression cloning and binding experiments

Sequences encoding mouse OMgp were amplified from Marathon-ready mouse cDNA (Clontech) and confirmed by sequencing analysis, before subcloning into the expression vector AP-5 (ref. 9) to express an AP-OM fusion protein tagged with both a polyhistidine and a Myc epitope. The resultant plasmid DNA was transfected into COS-7 cells and the secreted protein purified using nickel-agarose resins (Qiagen).

For expression cloning of OMgp-binding proteins, pools of 5,000 arrayed clones from a human brain cDNA library (Origene Technologies) were transfected into COS-7 cells, and cell-surface binding with AP-OM was performed as described previously^{4,13,17}. We isolated single NgR cDNA clones by subdividing the pools and sequencing analysis.

Recombinant proteins and viruses and co-precipitation

To express recombinant OMgp for functional assays, we subcloned the coding region of mouse OMgp (amino acids 23–392) into pSecTag B (Invitrogen) to express His-tagged OMgp protein (OM-His) in COS-7 cells. The expressed OM-His protein was purified using a nickel resin. To construct recombinant HSV, cDNAs for Flag-tagged NgR or β -galactosidase were inserted into the HSV amplicon HSV-PrpUC and packaged into the virus using helper 5dl1.2, as described previously²⁰. To produce recombinant Nogo-66 protein, the sequence of Nogo-66 was amplified from a human cDNA clone, KIAA0886 (Kazusa DNA Research Institute), and used to generate a construct expressing the AP-66 protein as described elsewhere⁶. Antibodies against Nogo-A and MAG were purchased from Alpha Diagnostics and R&D Systems, respectively.

In co-precipitation experiments, $2 \mu\text{g}$ GST or GST–NgR were first immobilized to glutathione-agarose beads and the beads were further incubated with or without $1 \mu\text{g}$ OM-His in the presence of $2 \mu\text{g}$ of AP or AP-66 at 4°C for 2 h. After extensive washing, bound proteins were resolved with SDS–PAGE and detected by western blotting.

Growth cone collapse and neurite outgrowth assays

Chick E13 DRG and E7 retina were isolated and cultured as described previously^{9,10}. DRG explants cultured overnight were used for growth cone collapse assays. In untreated control cultures, 80–85% of the growth cones were intact. To assess the effects of PI-PLC treatment, cultures were pre-incubated with 2 U ml^{-1} PI-PLC for 30 min before treatment with individual test proteins for another 30 min. To express NgR in E7 retinal ganglion neurons, we infected the explants with recombinant HSV for 24 h. After incubation with each test protein for 30 min, retinal explants were fixed in 4% paraformaldehyde and 15% sucrose. Infection of HSV–LacZ was detected by a standard β -galactosidase staining protocol²⁰. Flag–NgR expression was detected by incubating paraformaldehyde-fixed

cultures with the M2 anti-Flag antibody. Bound antibody was detected by incubation with AP-conjugated anti-rabbit immunoglobulin- γ (IgG) secondary antibody and reaction with nitroblue tetrazolium (NBT) and 5-bromo-4-chloro-3-indoxyl phosphate (BCIP)²². Growth cone collapse was quantified only in those positively stained for β -galactosidase or immunoreactive for the Flag epitope. The means $\pm$ s.e.m. from more than 400 growth cones per condition in three separate experiments are presented.

Neurite outgrowth assays were performed as described previously^{15,21}. Briefly, P7–9 rat CGNs were dissected and then plated at a density of 1×10^5 cells per well. Cells were cultured for 24 h before being fixed with 4% paraformaldehyde and stained with a neuronal-specific anti- β -tubulin antibody (Tuj-1, Covance). Neurite lengths were measured from at least 150 CGNs per condition, from duplicate wells per experiment, and from three independent experiments, and quantified as described previously²³.

Received 3 April; accepted 23 May 2002; doi:10.1038/nature00867.

Published online 16 June 2002.

1. Schwab, M. E. & Bartholdi, D. Degeneration and regeneration of axons in the lesioned spinal cord. *Physiol. Rev.* **76**, 319–370 (1996).
2. Horner, P. J. & Gage, F. H. Regenerating the damaged central nervous system. *Nature* **407**, 963–970 (2000).
3. McKerracher, L. *et al.* Identification of myelin-associated glycoprotein as a major myelin-derived inhibitor of neurite growth. *Neuron* **13**, 805–811 (1994).
4. Mukhopadhyay, G. *et al.* A novel role for myelin-associated glycoprotein as an inhibitor of axonal regeneration. *Neuron* **13**, 757–767 (1994).
5. Chen, M. S. *et al.* Nogo-A is a myelin-associated neurite outgrowth inhibitor and an antigen for monoclonal antibody IN-1. *Nature* **403**, 434–439 (2000).
6. GrandPré, T., Nakamura, F., Vartanian, T. & Strittmatter, S. M. Identification of the Nogo inhibitor of axon regeneration as a Reticulon protein. *Nature* **403**, 439–444 (2000).
7. Prinjha, R. *et al.* Inhibitor of neurite outgrowth in humans. *Nature* **403**, 383–384 (2000).
8. Tessier-Lavigne, M. & Goodman, C. S. Perspectives: neurobiology. Regeneration in the Nogo zone. *Science* **287**, 813–814 (2000).
9. Fournier, A. E., GrandPré, T. & Strittmatter, S. M. Identification of a receptor mediating Nogo-66 inhibition of axonal regeneration. *Nature* **409**, 341–346 (2001).
10. Luo, Y., Raible, D. & Raper, J. A. Collapsin: a protein in brain that induces the collapse and paralysis of neuronal growth cones. *Cell* **75**, 217–227 (1993).
11. Mikol, D. D. & Stefansson, K. A phosphatidylinositol-linked peanut agglutinin-binding glycoprotein in central nervous system myelin and on oligodendrocytes. *J. Cell Biol.* **106**, 1273–1279 (1988).
12. Mikol, D. D., Gulcher, J. R. & Stefansson, K. The oligodendrocyte-myelin glycoprotein belongs to a distinct family of proteins and contains the HNK-1 carbohydrate. *J. Cell Biol.* **110**, 471–479 (1990).
13. He, Z. & Tessier-Lavigne, M. Neuropilin is a receptor for the axonal chemorepellent Semaphorin III. *Cell* **90**, 739–751 (1997).
14. Habib, A. A. *et al.* Expression of the oligodendrocyte-myelin glycoprotein by neurons in the mouse central nervous system. *J. Neurochem.* **70**, 1704–1711 (1998).
15. Niederost, B. P., Zimmermann, D. R., Schwab, M. E. & Bandtlow, C. E. Bovine CNS myelin contains neurite growth-inhibitory activity associated with chondroitin sulfate proteoglycans. *J. Neurosci.* **19**, 8979–8989 (1999).
16. Spillmann, A. A., Bandtlow, C. E., Lottspeich, F., Keller, F. & Schwab, M. E. Identification and characterization of a bovine neurite growth inhibitor (bNI-220). *J. Biol. Chem.* **273**, 19283–19293 (1998).
17. Flanagan, J. G. & Cheng, H. J. Alkaline phosphatase fusion proteins for molecular characterization and cloning of receptors and their ligands. *Methods Enzymol.* **327**, 198–210 (2000).
18. Liu, B. P. & Strittmatter, S. M. Semaphorin-mediated axonal guidance via Rho-related G proteins. *Curr. Opin. Cell Biol.* **13**, 619–626 (2001).
19. Norton, W. T. & Poduslo, S. E. Myelination in rat brain: method of myelin isolation. *J. Neurochem.* **21**, 749–757 (1973).
20. Neve, R. L., Howe, J. R., Hong, S. & Kalb, R. G. Introduction of the glutamate receptor subunit 1 into motor neurons in vitro and in vivo using recombinant herpes simplex virus. *Neuroscience* **79**, 435–447 (1997).
21. Huang, D. W., McKerracher, L., Braun, P. E. & David, S. A therapeutic vaccine approach to stimulate axon regeneration in the adult mammalian spinal cord. *Neuron* **24**, 639–647 (1999).
22. Takahashi, T., Nakamura, F., Jin, Z., Kalb, R. G. & Strittmatter, S. M. Semaphorins A and E act as antagonists of neuropilin-1 and agonists of neuropilin-2 receptors. *Nature Neurosci.* **1**, 487–493 (1998).
23. Cohen-Cory, S. & Fraser, S. E. Effects of brain-derived neurotrophic factor on optic axon branching and remodeling in vivo. *Nature* **378**, 192–196 (1995).

Supplementary Information accompanies the paper on Nature's website (<http://www.nature.com/nature>).

Acknowledgements

We thank A. Habib and J. Flanagan for providing the anti-OMgp antibodies and the AP-5 plasmid, and M. Greenberg, X. He, D. Lowenstein and M. Tessier-Lavigne for reading the manuscript. This study was supported by the Alfred Sloan Foundation, the Burroughs Wellcome Fund, the EHLB Foundation, the International Spinal Research Trust, the John Merck Fund, the Klingenstein Fund, the Whitehall Foundation, and the National Institutes of Health (NIH). K.C.W. is a recipient of a Howard Hughes Predoctoral Fellowship. V.K. is supported by an NIH predoctoral training grant.

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to Z.H. (e-mail: zhiqiang.he@tch.harvard.edu).

Placental-specific IGF-II is a major modulator of placental and fetal growth

Miguel Constância*, Myriam Hemberger†, Jennifer Hughes*, Wendy Dean*, Anne Ferguson-Smith‡, Reinald Fundele§, Francesca Stewart*, Gavin Kelsey*, Abigail Fowden||, Colin Sibley¶ & Wolf Reik*

* Laboratory of Developmental Genetics and Imprinting, Developmental Genetics Programme, The Babraham Institute, Cambridge CB2 4AT, UK

† Department of Biochemistry and Molecular Biology, University of Calgary, Calgary, Alberta T2N 4N1, Canada

‡ Department of Anatomy, University of Cambridge, Cambridge CB2 3DY, UK
§ Max-Planck-Institut für Molekulare Genetik, D-14195 Berlin-Dahlem, Germany

|| The Physiological Laboratory, University of Cambridge, Cambridge CB2 3EG, UK

¶ Academic Unit of Child Health, The University of Manchester, St Mary's Hospital, Manchester M13 0JH, UK

Imprinted genes in mammals are expressed from only one of the parental chromosomes, and are crucial for placental development and fetal growth^{1–4}. The insulin-like growth factor II gene (*Igf2*) is paternally expressed in the fetus and placenta⁵. Here we show that deletion from the *Igf2* gene of a transcript (P0)^{6,7} specifically expressed in the labyrinthine trophoblast of the placenta leads to reduced growth of the placenta, followed several days later by fetal growth restriction. The fetal to placental weight ratio is thus increased in the absence of the P0 transcript. We show that passive permeability for nutrients of the mutant placenta is decreased, but that secondary active placental amino acid transport is initially upregulated, compensating for the decrease in passive permeability. Later the compensation fails and fetal growth restriction ensues. Our study provides experimental evidence for imprinted gene action in the placenta that directly controls the supply of maternal nutrients to the fetus, and supports the genetic conflict theory of imprinting⁸. We propose that the *Igf2* gene, and perhaps other imprinted genes, control both the placental supply of, and the genetic demand for, maternal nutrients to the mammalian fetus.

A substantial proportion of imprinted genes are involved in the control of fetal growth and placental development^{1–4}. Deletion of the paternally expressed *Igf2*, *Peg1/Mest*, *Peg3* or *Ins1/Ins2* genes results in intrauterine growth restriction^{5,9–11} (IUGR), whereas deletion of the maternally expressed *Igf2r* or *H19* genes, or over-expression of the *Igf2* gene, results in fetal overgrowth^{12–15}. The maternally expressed *Mash2* gene is important for the development of the spongiotrophoblast in the placenta¹⁶, and other imprinted genes are involved in the organization and behaviour of placental cell types¹⁷. The control of fetal growth by imprinted genes can be exerted at the level of cell proliferation and apoptosis in the fetus^{18,19}. Because of their effects on placental structure and physiology, imprinted genes could also control fetal growth by acting in the placenta to influence the supply of maternal nutrients to the fetus¹⁸. According to the genetic conflict hypothesis, paternally expressed genes acting in the placenta are predicted to extract more resources from the mother to enhance fetal growth, whereas maternally expressed genes are predicted to restrain fetal growth to conserve resources in the interest of the lifetime reproductive fitness of the mother⁸. This principal hypothesis in imprinting has been impossible to test because the major growth-controlling genes mentioned above are expressed both in the placenta and in the fetus. However, a recent study using chimaeras with IGF-II-deficient cells in the

trophoectoderm showed some effects on fetal growth, suggesting the possibility that placental IGF-II could be important for fetal growth²⁰.

We have previously described⁶ a new promoter (P0) in the mouse *Igf2* gene that leads to paternal expression of a transcript containing the *Igf2* coding exons 4–6 (P0 transcript) specifically in the labyrinthine trophoblast cells of the placenta (Fig. 1a, b). This provided a unique opportunity to reduce IGF-II action specifically in the placenta. We abolished the P0 transcript *in vivo* by deleting a 5-kilobase (kb) region spanning exon u2 of the mature P0 messenger RNA⁷ (Fig. 1a). Heterozygous placentae that inherited the P0 mutation from the paternal germ line (designated +/–) showed no detectable P0 mRNA, when using an exon u2-specific probe (Fig. 1b) or u1 probe (data not shown). The deletion thus has the same effect as a P0 promoter deletion. The relative contribution of the P0 transcript to overall levels of *Igf2* transcripts in placenta was evaluated by *in situ* hybridization with a full-length complementary DNA probe that detects all *Igf2* transcripts. A marked reduction in *Igf2* signal was observed specifically in the labyrinthine trophoblast of the mutant placentae (Fig. 1c). *Igf2* transcripts from other promoters were expressed appropriately and at wild-type levels in the mutant placentae and fetal tissues, and levels of IGF-II peptide in the fetal circulation were unaltered (data not shown). The deletion therefore eliminated specifically the P0 transcript.

Lack of the P0 transcript with paternal transmission of the deletion resulted primarily in placental growth restriction. Placental growth was impaired as early as embryonic day 12 (E12), when the mean weight of mutant placentae was 76% of wild-type placentae (average (in g) $\pm$ standard error: +/+, 0.030 ± 0.001 , $n = 27$; +/–, 0.023 ± 0.001 , $n = 25$; $P < 0.001$). The degree of growth deficiency of mutant placentae was relatively constant throughout mid- and late gestation (82%, 68% and 68% of wild-type at E14, E16 and E19, respectively, $P < 0.001$) (Fig. 2a). Notably, the degree and the time of onset of placental growth deficiency in P0 mutants was identical to that in *Igf2* null mice which lack the IGF-II peptide in all placental layers as well as in the fetus²¹ (our unpublished observations). These findings show that the labyrinthine trophoblast-specific P0 transcript is critical for sustaining normal growth of the placenta.

Notably, mutant fetuses were growth-restricted only towards the end of gestation (96% and 78% of wild type at E16 and E19, respectively) (Fig. 2b). Consequently, the fetal to placental weight ratio was significantly higher in the mutants compared with the wild type throughout gestation (Fig. 2c). At birth, P0 mutant pups were 69% of normal birth weight. This was followed by post-natal 'catch-up' growth, which was complete by three months of age (data not shown).

The placental and fetal growth kinetics in the P0 mutants suggest either that placental size becomes limiting for fetal growth only at later stages of gestation, or that at earlier stages the smaller placenta compensates by an increased efficiency of nutrient transfer. To address these hypotheses we investigated placental structure and transport capacity in mutant and wild-type mice.

Mutant placentae revealed the same morphological organization as wild type, with no obvious differences in cell morphology. Volume fraction determinations (that is, percentages of total placental volume) for decidua, spongiotrophoblast and labyrinthine layers showed that all areas were proportionately smaller (data not shown).

We measured unidirectional maternal–fetal transfer of ⁵¹Cr-EDTA and the non-metabolizable amino acid analogue ¹⁴C-methylaminoisobutyric acid (¹⁴C-MeAIB) as fetal accumulation of radioisotope at different time points after injection into the maternal circulation. ⁵¹Cr-EDTA is an inert hydrophilic molecule that provides a measure of the passive permeability of the placenta to hydrophilic solute²², whereas ¹⁴C-MeAIB is transported specifically by the system A amino acid transporter in the placenta²³ and was

used as an indicator of secondary active transport capacity. These measurements were made at E16, when the mutant placenta is smaller but the fetus is still about the same size as wild type, and at E19, when both are smaller than the wild type (Fig. 2).

At E16, ^{51}Cr -EDTA fetal accumulation, per gram of mutant placenta, was only 80% of that per gram of wild-type placenta ($P = 0.052$) (Fig. 3a). The ^{51}Cr -EDTA fetal accumulation per gram of mutant placenta was further reduced to 60% of the wild-type level at E19 ($P < 0.001$) (Fig. 3a). This shows that at E16, and more severely at E19, passive permeability of the placentae to hydrophilic solute, per gram, was reduced in P0 mutants. Coupled to the reduced total placental size, this defect would be predicted to severely restrict fetal growth at both gestational ages. Indeed, mutant fetuses received significantly less solute than wild-type ones at both time points (Fig. 3b).

In contrast, mutant placentae transferred more ^{14}C -MeAIB per gram compared with wild-type counterparts at E16 (155%, $P < 0.001$) (Fig. 3a). This shows that P0 mutant placentae actively transferred the solute with greater efficiency. Thus at E16 the mutant fetuses received the same actual amount of ^{14}C -MeAIB as wild type (consistent with their having the same body weight) (Fig. 3b). By E19 the increase in transfer per gram of placenta was much reduced (116%, $P < 0.05$) (Fig. 3a), and this combined with the smaller size of the placenta resulted in mutant fetuses receiving less of the solute than wild-type littermates (74%, $P < 0.001$) (Fig. 3b). Notably, this reduction was proportionate to the reduction in fetal weight (78%, Fig. 2b). This suggests that at the earlier stage of gestation, increased efficiency of secondary active transfer compensated for reduced placental size and passive permeability. At the later stage of gestation, the increase in efficiency of the P0 mutant

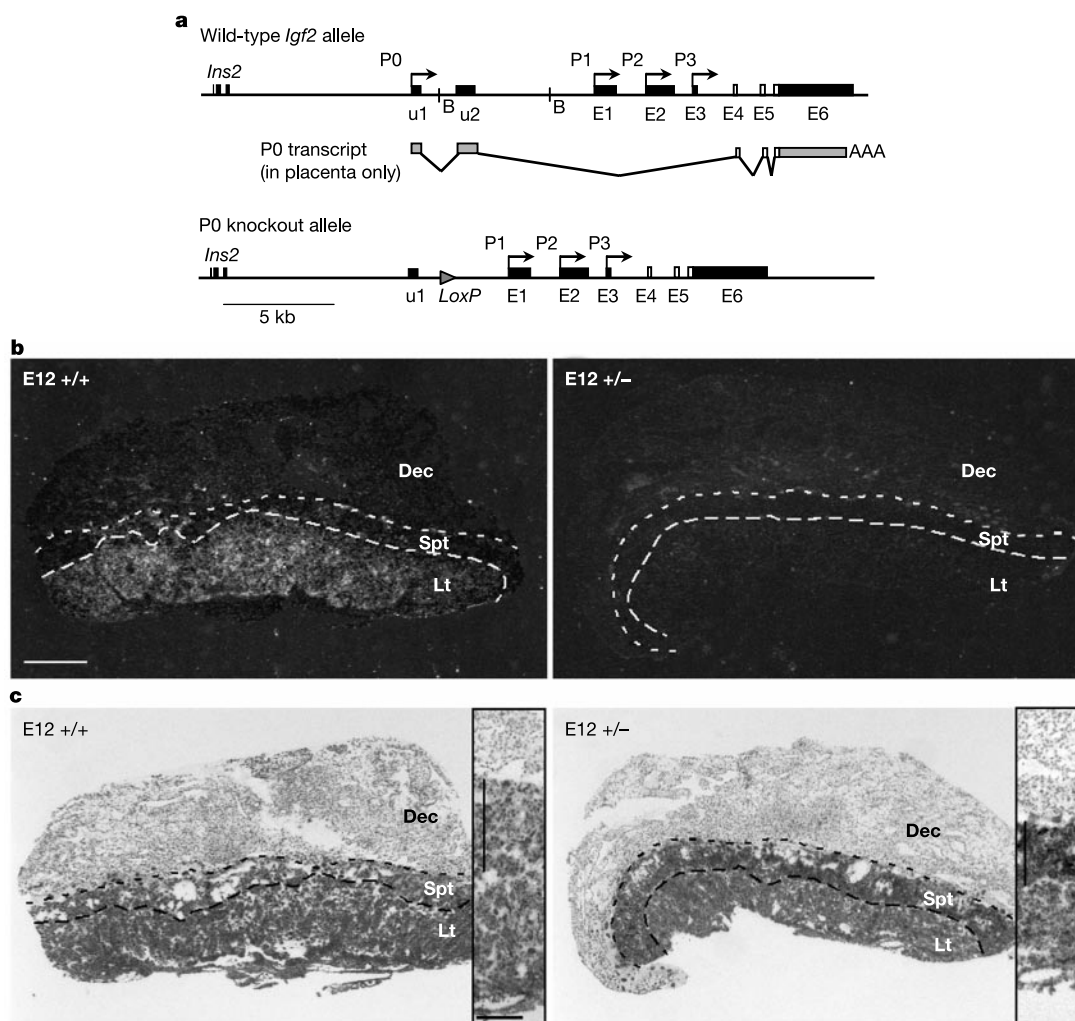


Figure 1 Characterization of the P0 mutation. **a**, Genomic structure of the mouse wild-type *Igf2* and knockout alleles. The peptide-coding region is in exons 4–6 (E4–E6, open boxes). Promoters P1–P3 are transcribed in fetal and extra-embryonic tissues. The P0 transcript is placenta-specific⁶. All transcripts are spliced onto exon 4, leading to the same peptide. A 5-kb *Bam*HI (B) genomic fragment was replaced by a *loxP* site in the P0 knockout allele⁷. The deletion, which spans exon u2 and abolishes P0 transcription, is 1.5 kb and 2.1 kb distant from P0 and P1 transcription start sites, respectively. **b**, mRNA *in situ* hybridization of sections from E12 wild-type (+/+) and P0 mutant (+/-) placentae, showing labyrinthine trophoblast-specific expression of the P0 transcript and its absence in the knockout placenta. An exon u2-specific antisense riboprobe was used. **c**, Relative contribution of the P0 transcript to overall *Igf2* transcript levels in placenta by

mRNA *in situ* hybridization. E12 wild-type (+/+) and P0 mutant (+/-) placentae were hybridized with a full-length *Igf2* cDNA antisense riboprobe (bright field view). Insets in each panel are higher magnification images. *Igf2* is expressed at a similar level in the spongiotrophoblast (vertical bar in insets) and labyrinthine trophoblast in the wild-type placenta (left panel). In the mutant P0 placenta (right panel), a marked reduction in expression is observed in the labyrinthine trophoblast when compared with the spongiotrophoblast (vertical bar in insets), due to the absence of the P0 transcript. Borders between labyrinthine trophoblast and spongiotrophoblast, and spongiotrophoblast and maternal decidua are shown as dashed lines. Dec, maternal decidua; Spt, spongiotrophoblast; Lt, labyrinthine trophoblast. Scale bars: 500 μm (main images); 200 μm (inset).

placentae is now insufficient to transfer enough amino acids to the fetus to meet its requirements for growth and this, together with the reduced passive permeability, results in fetal growth restriction. Importantly, the system A transporter has been shown to be a determinant of fetal growth²⁴.

The absence of the P0 transcript of the *Igf2* gene resulted in considerable deficiency in growth of the placenta, which was as great

as with complete lack of placental IGF-II, suggesting that the P0 transcript is the principal determinant of action of this growth factor in the placenta. This is perhaps surprising because the P0 transcript is estimated to constitute 10% of total *Igf2* transcripts in the placenta⁶; however, differential translatability of *Igf2* transcripts has been reported^{25,26}. This reduction in growth was associated with a decrease in passive permeability per gram of placenta. The mechanisms underlying this decreased permeability are unknown at present but may involve changes in placental exchange barrier paracellular pore dimensions. In association with the decreased permeability we found an increase in the fetal to placental weight ratio and an increase in MeAIB transfer per gram of placenta. The specific effector of the increased MeAIB transfer efficiency remains to be elucidated. We also need to determine whether other trophoblast transport systems are affected. However, this genetic model has demonstrated that upregulation of active transport may compensate for decreased size and diffusional capacity of the placenta, and that this mechanism is important in determining fetal growth rate. Genetic manipulation of functional properties of the placenta may also provide valuable models of 'fetal programming' of diabetes and heart disease later on in life^{27,28}.

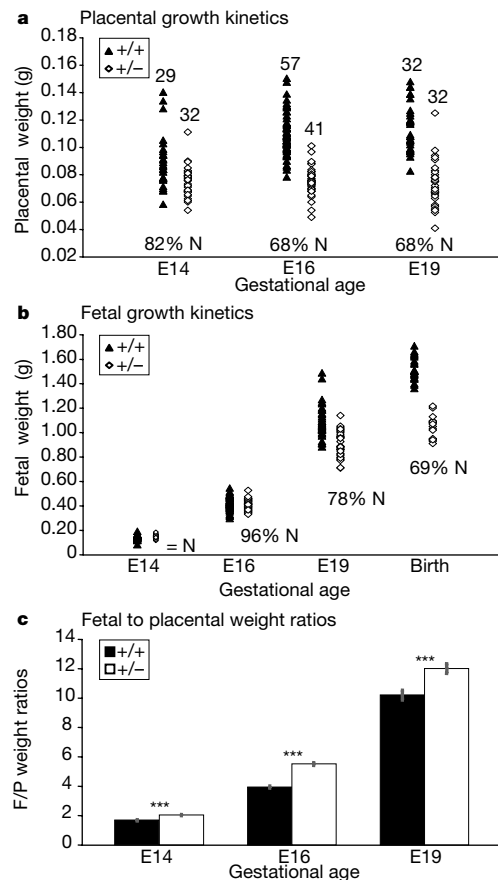


Figure 2 Growth kinetics during development. **a**, Placental weights of P0 mutant mice (+/-) and wild-type littermates (+/+) at the indicated gestational ages. Each symbol represents a single placenta (superimposed in some cases) with the total number of weighed placentae indicated at the top. The data were collected from 8, 13 and 8 litters (E14, E16 and E19, respectively). The degree of growth deficiency of P0 mutant placentae is shown as a percentage of normal placental weight (% N). The average placental weights (in g) $\pm$ standard error were: E14, +/+ = 0.091 $\pm$ 0.002, +/- = 0.075 $\pm$ 0.002, $P < 0.001$; E16, +/+ = 0.110 $\pm$ 0.002, +/- = 0.075 $\pm$ 0.002, $P < 0.001$; E19, +/+ = 0.113 $\pm$ 0.002, +/- = 0.077 $\pm$ 0.002, $P < 0.001$. **b**, Fetal weights of P0 mutants (+/-) and wild-type littermates (+/+) at the indicated gestational ages and at birth. The data are from the same litters as in **a**. The data at birth were collected from 6 litters (16 +/- and 24 +/+). At E14 and E12 (not shown) no differences in fetal weight were observed between P0 mutant and wild-type littermates. The degree of growth deficiency of P0 mutants is shown as a percentage of normal fetal weight (% N). The average weights (in g) $\pm$ standard error were: E14, +/+ = 0.150 $\pm$ 0.004, +/- = 0.149 $\pm$ 0.003, $P > 0.05$; E16, +/+ = 0.427 $\pm$ 0.006, +/- = 0.410 $\pm$ 0.008, $P < 0.05$; E19, +/+ = 1.138 $\pm$ 0.015, +/- = 0.886 $\pm$ 0.015, $P < 0.001$; birth, +/+ = 1.523 $\pm$ 0.019, +/- = 1.046 $\pm$ 0.023, $P < 0.001$. **c**, Fetal to placental weight ratios (F/P) were calculated for P0 mutant mice (+/-) and wild-type littermates (+/+). Fetal to placental ratio was significantly higher in mutants compared with the wild type throughout gestation (1.20, 1.40 and 1.18 at E14, E16 and E19, respectively; triple asterisk, $P < 0.001$). The average fetal to placental ratios $\pm$ standard error were: E14, +/+ = 1.682 $\pm$ 0.042, +/- = 2.043 $\pm$ 0.040; E16, +/+ = 3.940 $\pm$ 0.076, +/- = 5.521 $\pm$ 0.095; E19, +/+ = 10.213 $\pm$ 0.333, +/- = 12.021 $\pm$ 0.325.

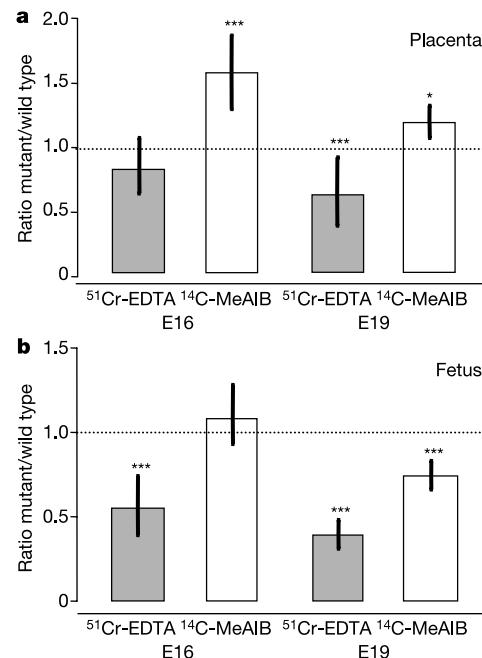


Figure 3 Placental transfer is altered in P0 mutants. **a**, Placental transfer of ⁵¹Cr-EDTA and ¹⁴C-MeAIB in P0 and wild-type conceptuses. The fetal accumulation of radioisotope is expressed relative to placental weight and plotted as a ratio of mutant to wild type at two gestational ages (E16, E19). This gives a relative measure of placental transfer of these solutes. ⁵¹Cr-EDTA ratio was lower than 1 at E16 (0.8) and E19 (0.60), indicating a decrease in passive permeability of the P0 mutant placentae. In contrast, ¹⁴C-MeAIB ratio was higher than 1 (E16, 1.55; E19, 1.16), showing that P0 placentae actively transferred the solute with greater efficiency. **b**, The fetal accumulation of ⁵¹Cr-EDTA and ¹⁴C-MeAIB is plotted as a ratio of mutant to wild-type (not expressed relative to fetal weight). This gives an absolute measure of the amount received by the fetus of the solutes. ⁵¹Cr-EDTA ratio was lower than 1 at E16 (0.55) and at E19 (0.39), indicating that the mutant fetuses received less solute than wild type. ¹⁴C-MeAIB ratio is approximately 1 at E16 (1.08), indicating that the upregulation of placental transport achieves the same amount of solute in mutant fetuses as in wild type, but is reduced at E19 (0.74), showing that the active transport at this stage no longer achieves the same amount of solute in the mutant fetuses as in wild-type ones. The data for the placental transport capacity assays were collected from 6 and 4 litters at E16 and E19, respectively (E16: +/+, $n = 25$, +/-, $n = 16$; E19: +/+, $n = 14$, +/-, $n = 18$). Bars indicate 95% confidence limits. Asterisk, $P < 0.05$; triple asterisk, $P < 0.001$.

Deletion of a placental-specific imprinted transcript results in fetal growth restriction, primarily through a decrease in total nutrient transfer across the placenta. This example of a morphologically normal but small placenta affecting fetal growth supports the genetic conflict theory of imprinting⁸, in which a placental-specific gene expressed from the paternal allele regulates the supply of nutritional resources to the fetus. On the other hand, fetal demand for nutrients is genetically regulated by the level of growth factors such as INS, IGF-I and IGF-II. Increasing fetal size therefore requires a higher level of demand (for example, higher fetal IGF-II) as well as a higher level of supply (by increasing, for example, placental surface area). Reduced fetal size can be the outcome of reduced supply (as in the P0 mutant described here) or of reduced demand (for example *Igf1* knockout, which reduces fetal but not placental size²¹). The mouse *Igf2* gene is remarkable in combining the control of both the supply and the genetic demand for maternal nutrients in a single gene.

Note added in proof: We have repeated the ⁵¹Cr-EDTA measurements at E16 in a larger experimental series (7 litters: 29 +/+ and 30 +/-) and observed that the fetal accumulation, per gram of mutant placenta, was 67% of that per gram of wild-type placenta ($P < 0.05$). The decrease in passive permeability of the P0 placenta is therefore significant at E16, as well as E19 as shown in Fig. 3a. □

Methods

Growth kinetic studies

Details on the generation of the P0 knockout allele are provided elsewhere⁷. The mutant P0 allele was bred into a C57BL/6J genetic background for analysis. For the growth kinetics studies, pregnant females were dissected at various embryonic stages (embryonic day 1 (E1) was defined as the day of vaginal plug detection). Wet weights of placenta and whole fetuses were recorded and mutant animals were distinguished from wild types with a Southern blotting-based assay, as shown before⁷.

mRNA *in situ* hybridization

Placentae were fixed overnight at 0 °C in Serra's fixative, dehydrated, and embedded in paraffin wax according to standard protocols. Seven-micrometre sections were processed and hybridized with UTP ³⁵S-labelled RNA probes as described²⁹. A 0.9-kb *SacI* genomic fragment comprising the mouse exon u2 and a 1.2-kb full-length *Igf2* mouse cDNA were used to generate sense and antisense RNA probes. *Igf2* transcript levels were determined by northern blotting as previously described⁷, and IGF-II peptide in serum extracts was measured by radio-immunoassay according to ref. 30, using an anti-human IGF-II rabbit polyclonal antibody (Gropep).

Morphometric analysis

Placentae were dissected at E18 of gestation, fixed in Serra's fixative and processed for routine paraffin histology. Sections were incubated with a peroxidase-conjugated lectin from *Bandeiraea simplicifolia* (BS-I B4, Sigma), and counterstained with haematoxylin. Three sections at distances of 35–50 µm were analysed for each placenta, providing a random sample of vertical sections. The volume fractions were determined by point counting for decidua, spongiotrophoblast and labyrinthine trophoblast. Morphometric analysis was carried out in three mutant and six wild-type placentae of the same litter.

Analysis of unidirectional maternal–fetal transfer

Radiolabelled ⁵¹Cr-EDTA (25 µCi, NEN) and ¹⁴C-methyl-aminoisobutyric acid (MeAIB) (3.5 µCi, NEN) in 100 µl phosphate buffered saline was injected into the jugular vein of C57BL/6J female mice bred with heterozygous P0 males, at E16 and E19 of pregnancy. Fetuses were killed at known times between 1 and 5 min after injection of radioisotopes (in preliminary experiments we determined that there was minimal backflux of radioisotope from fetus to mother up to 5 min). Fetuses and placentae were weighed and a small section of tail removed for genotyping. For lysis of the fetus and placental tissue, the samples were incubated overnight at 55 °C in Biosol (National Diagnostics). Fractions were then either added to liquid scintillation fluid for β-counting or to appropriate tubes for γ-counting. Radioactive counts in each fetus were then used to calculate the amount of radioisotope transferred per gram of placenta or per whole embryo. Average values for wild-type and mutant fetuses within a litter were then calculated and expressed as a ratio of mutant to wild type for that litter. These values could then be used to calculate a mean for all litters at E16 and E19.

Statistical analysis

All data were analysed by means of two-way analyses of variance with 'litters' and 'genotype' as the two factors. For data representing radioactive counts, a logarithmic transformation was carried out before analysis. The summaries were then presented as ratios, together with 95% confidence limits. Data representing weights of the offspring were analysed on the original scale.

Received 22 January; accepted 3 May 2002; doi:10.1038/nature00819.

- Reik, W. & Walter, J. Genomic imprinting: parental influence on the genome. *Nature Rev. Genet.* **2**, 21–32 (2001).
- Ferguson-Smith, A. C. & Surani, M. A. Imprinting and the epigenetic asymmetry between parental genomes. *Science* **293**, 1086–1089 (2001).
- Slueteels, F., Barlow, D. P. & Lyle, R. The uniqueness of the imprinting mechanism. *Curr. Opin. Genet. Dev.* **10**, 229–233 (2000).
- Tilghman, S. M. The sins of the fathers and mothers: genomic imprinting in mammalian development. *Cell* **96**, 185–193 (1999).
- DeChiara, T. M., Robertson, E. J. & Efstratiadis, A. Parental imprinting of the mouse insulin-like growth factor II gene. *Cell* **64**, 849–859 (1991).
- Moore, T. *et al.* Multiple imprinted sense and antisense transcripts, differential methylation and tandem repeats in a putative imprinting control region upstream of mouse *Igf2*. *Proc. Natl Acad. Sci. USA* **94**, 12509–12514 (1997).
- Constância, M. *et al.* Deletion of a silencer element in *Igf2* results in loss of imprinting independent of *H19*. *Nature Genet.* **26**, 203–206 (2000).
- Moore, T. & Haig, D. Genomic imprinting in mammalian development: a parental tug-of-war. *Trends Genet.* **7**, 45–49 (1991).
- Lefebvre, L. *et al.* Abnormal maternal behaviour and growth retardation associated with loss of the imprinted gene *Mest*. *Nature Genet.* **20**, 163–169 (1998).
- Li, L. *et al.* Regulation of maternal behaviour and offspring growth by paternally expressed *Peg3*. *Science* **284**, 330–333 (1999).
- Duvillie, B. *et al.* Phenotypic alterations in insulin-deficient mutant mice. *Proc. Natl Acad. Sci. USA* **94**, 5137–5140 (1997).
- Wang, Z.-Q., Fung, M. R., Barlow, D. P. & Wagner, E. F. Regulation of embryonic growth and lysosomal targeting by the imprinted *Igf2r/Mpr* gene. *Nature* **372**, 464–467 (1994).
- Lau, M. M. *et al.* Loss of imprinted IGF2/cation-independent mannose 6-phosphate receptor results in fetal overgrowth and perinatal lethality. *Genes Dev.* **8**, 2953–2963 (1994).
- Leighton, P. A., Ingram, R. S., Eggenschwiler, J., Efstratiadis, A. & Tilghman, S. M. Disruption of imprinting caused by deletion of the *H19* region in mice. *Nature* **375**, 34–39 (1995).
- Sun, F. L., Dean, W. L., Kelsey, G., Allen, N. D. & Reik, W. Transactivation of *Igf2* in a mouse model for Beckwith–Wiedemann syndrome. *Nature* **389**, 809–815 (1997).
- Guillemot, F. *et al.* Genomic imprinting of *Mash2*, a mouse gene required for trophoblast development. *Nature Genet.* **9**, 235–242 (1995).
- Georgiades, P., Watkins, M., Burton, G. J. & Ferguson-Smith, A. C. Roles for genomic imprinting and the zygotic genome in placental development. *Proc. Natl Acad. Sci. USA* **98**, 4522–4527 (2001).
- Reik, W., Davies, K., Dean, W., Kelsey, G. & Constância, M. Imprinted genes and the coordination of fetal and postnatal growth in mammals. *Novartis Found. Symp.* **237**, 19–31 (2001).
- Burns, J. L. & Hassan, A. B. Cell survival and proliferation are modified by insulin-like growth factor 2 between days 9 and 10 of mouse gestation. *Development* **128**, 3819–3830 (2001).
- Gardner, R. L., Squire, S., Zaina, S., Hills, S. & Graham, C. F. Insulin-like growth factor-2 regulation of conceptus composition: effects of the trophoctoderm and inner cell mass genotypes in the mouse. *Biol. Reprod.* **60**, 190–195 (1999).
- Baker, J., Liu, J.-P., Robertson, E. J. & Efstratiadis, A. Role of insulin-like growth factors in embryonic and postnatal growth. *Cell* **75**, 73–82 (1993).
- Atkinson, D. E., Robinson, N. R. & Sibley, C. P. Development of the passive permeability characteristics of the rat placenta during the last third of gestation. *Am. J. Physiol.* **261**, R1461–R1464 (1991).
- Johnson, L. W. & Smith, C. H. Neutral amino acid transport systems of microvillous membrane of human placenta. *Am. J. Physiol.* **251**, C773–C780 (1988).
- Cramer, S., Beveridge, M., Kilberg, M. & Novak, D. Physiological importance of system A-mediated amino acid transport to rat fetal development. *Am. J. Physiol. Cell. Physiol.* **282**, C153–C160 (2002).
- Newell, S., Ward, A. & Graham, C. Discriminating translation of insulin-like growth factor-II (IGF-II) during mouse embryogenesis. *Mol. Reprod. Dev.* **39**, 249–258 (1994).
- Nielsen, J. *et al.* A family of insulin-like growth factor II mRNA-binding proteins represses translation in late development. *Mol. Cell Biol.* **19**, 1262–1270 (1999).
- Godfrey, K. M. & Barker, D. J. Fetal programming and adult health. *Public Health Nutr.* **4**, 611–624 (2001).
- Ozanne, S. E. & Hales, C. N. The long-term consequences of intra-uterine protein malnutrition for glucose metabolism. *Proc. Nutr. Soc.* **58**, 615–619 (1999).
- Adam, G. L., Cui, H., Miller, S. J., Flam, F. & Ohlsson, R. Allele-specific *in situ* hybridization (ASISH) analysis: a novel technique which resolves differential allelic usage of *H19* within the same cell lineage during human placental development. *Development* **122**, 839–847 (1996).
- Hill, D. J. Relative abundance and molecular size of immunoreactive insulin-like growth factors I and II in human fetal tissues. *Early Hum. Dev.* **21**, 49–58 (1990).

Acknowledgements

We dedicate this paper to Francesca Stewart who died in December 2000 and who was instrumental in initiating this collaborative work. We thank D. Hill for help with IGF-II serum assays, E. Walters for help with the statistical analyses, T. Moore for discussions, and C. Graham for suggestions on the manuscript. Our work is supported by the Medical Research Council (MRC) and the Biotechnology and Biological Sciences Research Council (BBSRC). G.K. is a Senior Fellow of the MRC.

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to M.C. (e-mail: miguel.constancia@bbsrc.ac.uk) or W.R. (e-mail: wolf.reik@bbsrc.ac.uk).

Mutations of the *BRAF* gene in human cancer

Helen Davies^{1,2}, Graham R. Bignell^{1,2}, Charles Cox^{1,2}, Philip Stephens^{1,2}, Sarah Edkins¹, Sheila Clegg¹, Jon Teague¹, Hayley Woffendin¹, Matthew J. Garnett³, William Bottomley¹, Neil Davis¹, Ed Dicks¹, Rebecca Ewing¹, Yvonne Floyd¹, Kristian Gray¹, Sarah Hall¹, Rachel Hawes¹, Jaime Hughes¹, Vivian Kosmidou¹, Andrew Menzies¹, Catherine Mould¹, Adrian Parker¹, Claire Stevens¹, Stephen Watt¹, Steven Hooper³, Rebecca Wilson³, Hiran Jayatilake⁴, Barry A. Gusterson⁵, Colin Cooper⁶, Janet Shipley⁶, Darren Hargrave⁷, Katherine Pritchard-Jones⁷, Norman Maitland⁸, Georgia Chenevix-Trench⁹, Gregory J. Riggins¹⁰, Darell D. Bigner¹⁰, Giuseppe Palmieri¹¹, Antonio Cossu¹², Adrienne Flanagan¹³, Andrew Nicholson¹⁴, Judy W. C. Ho¹⁵, Suet Y. Leung¹⁶, Siu T. Yuen¹⁶, Barbara L. Weber¹⁷, Hilliard F. Seigler¹⁸, Timothy L. Darrow¹⁸, Hugh Paterson³, Richard Marais³, Christopher J. Marshall³, Richard Wooster^{1,6}, Michael R. Stratton^{1,4} & P. Andrew Futreal¹

¹Cancer Genome Project, The Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, CB10 1SA, UK

³Cancer Research UK Centre for Cell and Molecular Biology, Chester Beatty Labs, Institute of Cancer Research, London SW3 6JB, UK

⁴Section of Cancer Genetics; ⁵Section of Molecular Carcinogenesis; and ⁷Section of Paediatrics, Institute of Cancer Research, Sutton, Surrey SM2 5NG, UK

⁵Department of Pathology, Western Infirmary, University of Glasgow, S11 6NT, UK

⁸Department of Biology, YCR Cancer Research Unit, University of York, York YO10 5YW, UK

⁹Queensland Institute of Medical Research, RBH Post Office Herston, Queensland 4029, Australia

¹⁰Department of Pathology, and ¹⁸Department of Surgery, Duke University Medical Centre, Durham, North Carolina 27710, USA

¹¹Institute of Molecular Genetics, C.N.R., Loc. Tramariglio, Alghero 07040, Italy

¹²Department of Pathology, University of Sassari, Azienda USL1, Sassari 07100, Italy

¹³Royal Free & University College Medical School, London WC1E 6JJ, UK

¹⁴Royal Brompton Hospital, London SW3 6NP, UK

¹⁵Department of Surgery, and ¹⁶Department of Pathology, The University of Hong Kong, Queen Mary Hospital, Hong Kong

¹⁷Abramson Family Cancer Research Institute, University of Pennsylvania Cancer Center, Philadelphia, Pennsylvania 19104, USA

²These authors contributed equally to this work

Cancers arise owing to the accumulation of mutations in critical genes that alter normal programmes of cell proliferation, differentiation and death. As the first stage of a systematic genome-wide screen for these genes, we have prioritized for analysis signalling pathways in which at least one gene is mutated in human cancer. The RAS–RAF–MEK–ERK–MAP kinase pathway mediates cellular responses to growth signals¹. RAS is mutated to an oncogenic form in about 15% of human cancer. The three *RAF* genes code for cytoplasmic serine/threonine kinases that are regulated by binding RAS^{1–3}. Here we report *BRAF* somatic missense mutations in 66% of malignant melanomas and at lower frequency in a wide range of human cancers. All mutations are within the kinase domain, with a single substitution (V599E) accounting for 80%. Mutated *BRAF* proteins have elevated kinase activity and are transforming in NIH3T3 cells. Furthermore, RAS function is not required for the growth of cancer cell lines with the V599E mutation. As *BRAF* is a serine/threonine kinase that is commonly activated by somatic point mutation in human cancer, it may provide new therapeutic opportunities in malignant melanoma.

Genomic DNA from 15 cancer cell lines (6 breast cancers, 1 small-cell lung cancer (SCLC), 6 non-small-cell lung cancers (NSCLC), 1 mesothelioma, 1 melanoma) and the corresponding matched lymphoblastoid cell lines from the same individuals were screened for sequence variants through the coding exons and intron–exon junctions of the *BRAF* gene using a capillary-based modified heteroduplex method followed by direct sequencing of polymerase chain reaction products. (Exon 1, containing 135 base pairs (bp) of coding sequence, failed to amplify despite the use of five different primer sets.) Three single-base substitutions were detected. Two were in *BRAF* exon 15: T1796A leading to a substitution of valine by glutamic acid at position 599 (V599E) in the melanoma cell line Colo-829, and C1786G leading to L596V in the NSCLC cell line NCI-H2087 (Fig. 1). A further mutation was found in exon 11: G1403C leading to G468A in the NSCLC cell line NCI-H1395. None of the three changes were present in the lymphoblastoid cell lines from the same individuals, indicating that the variants were somatically acquired mutations.

phoblastoid cell lines from the same individuals were screened for sequence variants through the coding exons and intron–exon junctions of the *BRAF* gene using a capillary-based modified heteroduplex method followed by direct sequencing of polymerase chain reaction products. (Exon 1, containing 135 base pairs (bp) of coding sequence, failed to amplify despite the use of five different primer sets.) Three single-base substitutions were detected. Two were in *BRAF* exon 15: T1796A leading to a substitution of valine by glutamic acid at position 599 (V599E) in the melanoma cell line Colo-829, and C1786G leading to L596V in the NSCLC cell line NCI-H2087 (Fig. 1). A further mutation was found in exon 11: G1403C leading to G468A in the NSCLC cell line NCI-H1395. None of the three changes were present in the lymphoblastoid cell lines from the same individuals, indicating that the variants were somatically acquired mutations.

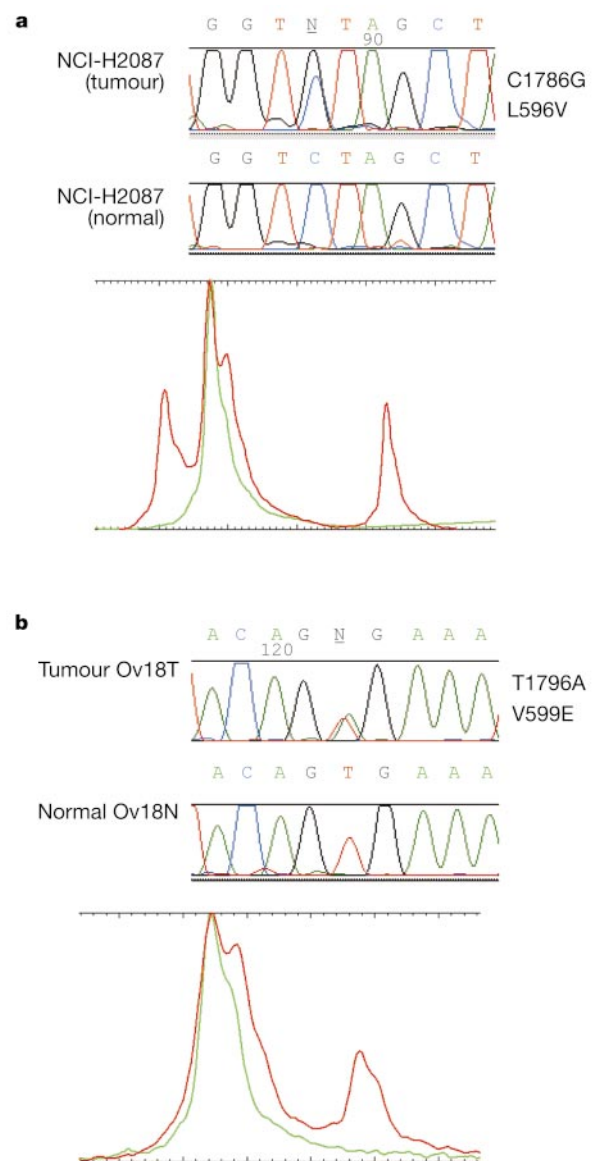


Figure 1 Mutations in the *BRAF* gene. Sequence electropherograms and corresponding comparisons between heteroduplex traces from normal (green) and cancer (red) DNAs from the same individuals. The heteroduplex trace comparisons are generated using proprietary software (see Methods). For each example (NCI-H2087 cell line (a) and ovarian neoplasm Ov18T (b)) the cancer trace shows additional peaks and/or differently shaped peaks compared with the normal trace.

letters to nature

To evaluate further the pattern of somatic mutations in *BRAF*, we screened the coding sequence and intron–exon junctions for mutations in genomic DNA from an additional 530 cancer cell lines (Table 1). Presumptive *BRAF* mutations (excluding variants of unknown significance and germline polymorphisms that were found in 341 normal tissue DNAs) were identified in 43 cancer cell lines including 20 of 34 (59%) melanomas, 7 of 40 (18%) colorectal cancers, 4 of 38 (11%) gliomas, 4 of 131 (3%) lung cancers (all four were adenocarcinomas from a total of 35), 5 of 59 (9%) sarcomas, 1 of 26 (4%) ovarian carcinomas, 1 of 45 (2%) breast cancers and 1 of 7 (14%) liver cancers. Mutations were not found in cancer cell lines derived from 29 neuroblastomas, 10 bladder cancers, 53 leukaemias and lymphomas, 11 cervical carcinomas, 11 renal cell carcinomas, 3 pancreatic carcinomas, 3 prostate carcinomas, 6 gastric carcinomas, 7 testicular carcinomas, 3 uterine carcinomas and 29 other cancers.

All 43 probable oncogenic *BRAF* somatic mutations found in the cancer cell lines were in exons 11 and 15 (Table 1). Accordingly, these two exons were screened for mutations in genomic DNA from 378 primary human cancers and short-term cultures (STC, less than passage 15). *BRAF* mutations were detected in 28 primary cancers/STCs, including 6 of 9 primary melanomas, 12 of 15 melanoma STCs, 4 of 33 colorectal carcinomas, 5 of 35 ovarian neoplasms (see Fig. 1 for example), and 1 of 182 sarcomas. Mutations were not detected in 33 breast cancers, 15 gliomas, 23 prostate cancers, 14 lung cancers, or 19 head and neck squamous cell carcinomas. Ten of the 35 ovarian tumours examined were classified as borderline (low malignant potential) lesions and 4 of 5 *BRAF* mutations found in ovarian neoplasms were in this subcategory. The single primary sarcoma in which a *BRAF* mutation was found was classified as a malignant fibrous histiocytoma.

Although *BRAF* mutations are found in a wide range of cancers, there is a trend towards the occurrence of mutations in cancer types in which a substantial proportion of cases are known to harbour *RAS* mutations (for example, malignant melanoma, colorectal cancer and borderline ovarian cancers^{4–6}). The apparent association between the presence of *BRAF* and *RAS* mutations in similar cancer

types suggests that activation of the RAS–RAF–mitogen-activated protein (MAP)–kinase kinase (MEK)–extracellular signal-regulated kinase (ERK)–MAP kinase pathway can be achieved by mutation at various levels in the pathway and that the pathway is activated in a substantial proportion of cases in these cancer types.

The highest frequency of *BRAF* mutations is in malignant melanoma (Table 1). This does not seem to be related to the effects of ultraviolet light, the only known environmental risk factor for this disease. The T → A change at nucleotide 1796, which accounts for 35 of 38 (92%) of *BRAF* mutations in melanoma (Table 1), is distinct from the CC → TT or C → T changes associated with pyrimidine dimer formation following exposure to ultraviolet light—these changes are commonly found, for example, in the *TP53* gene in non-melanoma skin cancers⁷.

The high frequency of mutation in melanoma may be related to features of melanocyte biology. α -melanocyte stimulating-hormone and other related proopiomelanocortin-derived peptides are crucial regulators in melanocyte biology. α -melanocyte stimulating-hormone and proopiomelanocortin-derived peptides bind to the melanocortin receptor 1, upregulating cyclic AMP, leading to increased proliferation and melanogenesis in response to UVB radiation⁸. This cAMP-dependent signalling cascade activates *BRAF* and subsequently ERK⁹. That a principal melanocyte-specific signalling pathway controlling proliferation and differentiation operates through activation of *BRAF* and that this gene is mutated in most melanomas suggests a possible explanation for the high frequency of *BRAF* mutation in melanomas relative to other cancer types.

Our analysis reveals mutations in two regions of the *BRAF* kinase domain. Mutations were very similarly distributed in cancer cell lines and primary cancers. A total of 89% of mutations are within or immediately adjacent to the activation segment, a region of 10–30 amino acids bounded by almost invariant DFG and APE motifs¹⁰. Acidic substitutions at a single amino acid residue (usually V599E and one instance of V599D) account for 92% of activation segment mutations with five further mutations altering residues E585, F594, G595 and L596 (Table 1). These residues are identical at the

Table 1 *BRAF* mutations in human cancer

<i>BRAF</i> mutations		Cancer cell lines								Primary tumours						Total
		(1) Mel.	(2) Colo. ca.	(3) Glioma	(4) Lung ca.	(5) Sarcoma	(6) Breast	(7) Ovarian	(8) Other	(1) Mel. STC	(2) Mel.	(3) Colo. ca.	(4) Ovarian*	(5) Sarcoma	(6) Other†	
G1388A	G463E							1								1
G1388T	G463V		1													1
G1394C	G465A									1						1
G1394A	G465E										1					1
G1394T	G465V				1											1
G1403C	G468A				2											2
G1403A	G468E											1				1
G1753A	E585K												1			1
T1782G	F594L											1				1
G1783C	G595R		1													1
C1786G	L596V				1											1
T1787G	L596R												1			1
T1796A	V599E	19	5	4		5	1		1	11	5	2	3	1	0	57
TG1796-97AT	V599D	1														1
	Total	20	7	4	4	5	1	1	1	12	6	4	5	1	0	71
No. samples screened		34	40	38	131	59	45	26	172	15	9	33	35	182	104	923
Per cent		59%	18%	11%	3%	9%	2%	4%	0.6%	80%	67%	12%	14%	0.5%	0%	8%

Amino acid residues are grouped in blocks. Three further *BRAF* coding sequence variants were identified (G2041A R681Q in the HEC1A endometrial cancer cell line, T974C I325T in the ZR-75-30 breast cancer cell line, and C2180T A727V in the H33AJ-JA1 T-ALL cell line). These were not present in 341 control DNAs. Lane numbers (in parentheses) are provided for convenience. Mel., melanoma; Colo. ca., colorectal cancer; Mel. STC, melanoma short-term culture.

*Four out of ten LMP (low malignant potential); 1 out of 25 malignant epithelial.

†Glioma (*n* = 15), breast cancer (*n* = 33), prostate cancer (*n* = 23), HNSCC (head and neck squamous cell carcinoma) (*n* = 19), lung cancer (*n* = 14).

equivalent positions in human *RAF1* and *ARAF1* and are conserved in all three *RAF* genes through evolution (with the exception of the *BRAF* V599 residue in the *Drosophila Raf* homologue; Fig. 2). A total of 11% of mutations are in the glycine residues (G463, G465 and G468) of the GXGXXG motif within the glycine-rich loop of the kinase domain. This motif is highly conserved in protein kinases and other proteins that bind mono- and dinucleotides (Fig. 2). The first glycine is present in about 95% of all kinases, the second in more than 99% of all kinases and the third in about 85% of kinases¹¹. From structural studies, it has been shown that this region forms a loop that anchors the β - and γ -phosphates of ATP and may orientate ATP for catalysis¹⁰. Previously described mutants of these glycine residues in other proteins have resulted in reduction of kinase activity^{12–14}. In the GTP-binding RAS proteins, however, mutation of G12 within the GXGXXG motif is still compatible with nucleotide binding and results in RAS proteins with transforming activity.

To characterize the biological effects of these mutations, we examined the V599E mutation—because it was the most common—together with one other mutant chosen from the activation segment (L596V) and two mutants of the G loop (G463V, G468A). Myc-epitope-tagged versions of complementary DNAs containing these mutations were transiently expressed in COS cells, immunoprecipitated using the Myc-tag and examined in a kinase cascade assay using bacterially produced glutathione *S*-transferase (GST)–MEK, GST–ERK and myelin basic protein (MBP) as sequential substrates^{15,16}. All four mutants had elevated basal kinase activity compared with wild-type BRAF (WT BRAF, Fig. 3a); however, the basal activity of G468A BRAF and V599E BRAF was substantially higher (about 12.5- and 10.7-fold that of WT BRAF, respectively), whereas the activation of G463V BRAF and L596V BRAF was more modest (about 2- and 5.7-fold, respectively). All four BRAF proteins also

stimulated the activity of endogenous ERK when expressed in COS cells as determined by a phospho-specific antibody that only binds to the dually phosphorylated activated form of ERK1/2 (Fig. 3b). Consistent with their *in vitro* activities, G468A BRAF and V599E BRAF stimulated phosphorylation of endogenous ERK1/2 more strongly than G463V BRAF or L596V BRAF. The data demonstrate that these mutants are active *in vitro* and stimulate the activity of the pathway *in vivo* to different degrees. Finally, all four mutants were also stimulated by G12V HRAS, although the fold activation for each of the mutants is reduced compared with WT BRAF (see Fig. 3a). L596V BRAF (about 4.5-fold stimulation by G12V HRAS) was stimulated more strongly than the other mutants tested (between 2- and 2.5-fold stimulation) (Fig. 3a).

The ability of the kinase-activated BRAF mutants to induce transformation was examined by transfection of the epitope-tagged cDNA constructs into NIH3T3 cells to assay focus-forming ability. WT BRAF transformed cells at very low efficiency (0.0013 foci per ng DNA); however, G463V BRAF, G468A BRAF, L596V BRAF and V599E BRAF transformed NIH3T3 cells 70–138 times (0.09–0.18 foci per ng DNA) more efficiently than WT BRAF (Table 2). Inhibiting the kinase activity by substituting alanine for aspartic acid at position 593 within the conserved DFG motif abrogated transform-

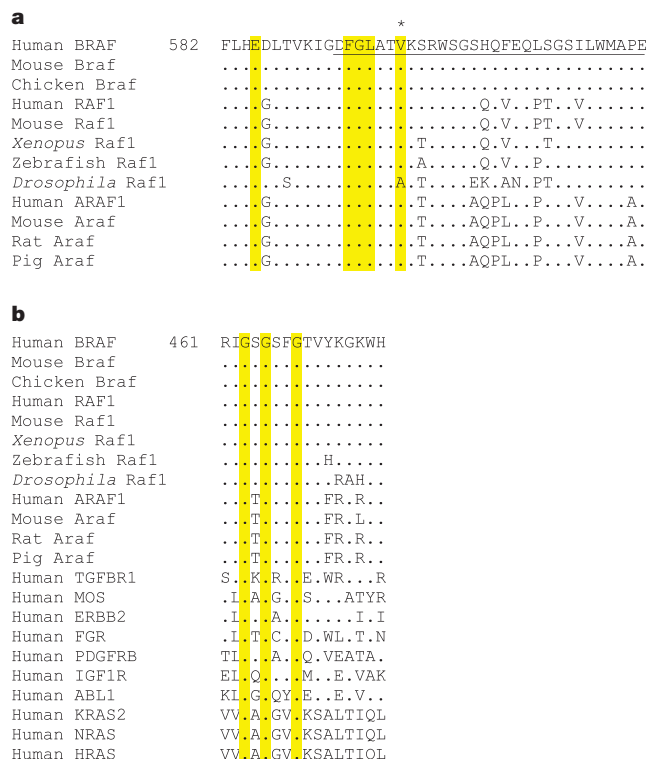


Figure 2 Sequence conservation and mutations in the *BRAF* activation segment and G loop. **a**, **b**, Conservation of amino acid sequence for the activation segment (**a**) and G loop (**b**). The positions of mutations are indicated by yellow shading; V599 is denoted by an asterisk. The activation segment is underlined.

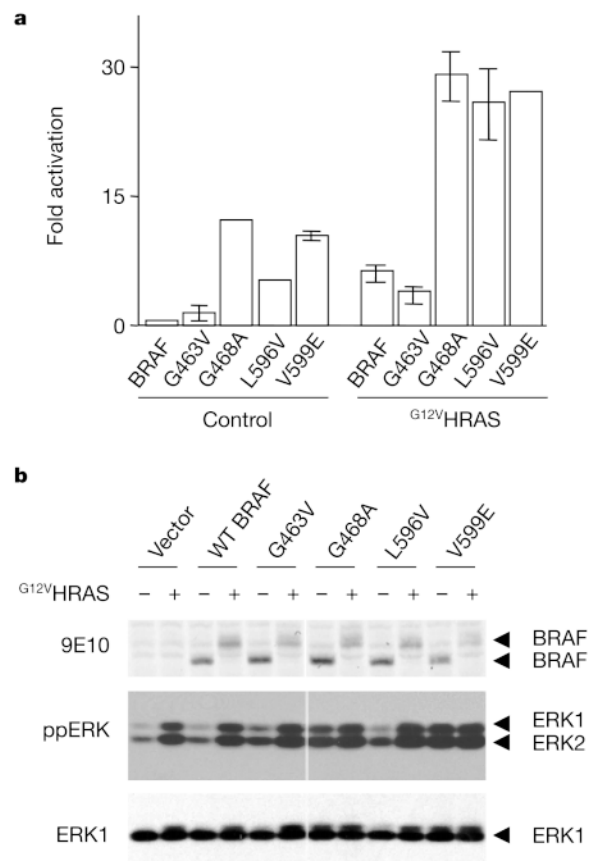


Figure 3 BRAF and ERK activation. **a**, BRAF kinase activity. Myc-epitope-tagged versions of BRAF and the various mutants were expressed in COS cells alone or in combination with G12V HRAS as indicated. The activity of the BRAF proteins in cell extracts was examined using the Raf kinase cascade assay. Each sample was assayed in triplicate and error bars are used to indicate the standard deviations from the mean. Absence of error bars indicates less than 3% error. Similar results were obtained in two independent transfections. **b**, ERK phosphorylation. Samples prepared as in **a** were examined for the presence of phosphorylated, active ERK1/2 (ppERK, middle panel) or total ERK (ERK, bottom panel). Similar results were obtained from extracts from two independent transfection assays. The top panel shows BRAF protein expression in transfectants.

Table 2 Transforming activity of BRAF mutants

Allele	Transformed foci per μ g DNA	Fold increase over wild-type BRAF
WT BRAF	1.3	—
V599E	180	138 $\times$
DAVE	0	—
L596V	90	70 $\times$
DALV	0	—
G463V	130	100 $\times$
G468A	90	69 $\times$
G12V HRAS	12,000	9,200 $\times$

NIH3T3 cells were transfected as described in Methods. Transformed foci contained cells like Ras- or Raf1-transformed cells—which are refractile and frequently bipolar—and often contained the giant cells typical of RAS or RAF1 transformation. DAVE and DALV are kinase-inactive versions of V599E and L596V, respectively, in which D593 of the conserved DFG motif is replaced by alanine to generate a kinase-dead variant.

ing activity, showing that this was dependent on kinase activity. Despite the higher kinase activity of V599E BRAF compared with WT BRAF activated by G12V HRAS, V599E BRAF has 50-fold lower transforming activity than G12V HRAS in this assay. This may reflect the fact that whereas RAS signals to a number of effector molecules¹⁸, RAF proteins may signal predominantly through the ERK–MAPK pathway (although other effectors have been proposed^{2,3,18}). In addition to focus assays, NIH3T3 cells transfected with G463V BRAF and G468A BRAF were assayed for tumorigenicity in nude mice. 10⁶ transfected cells pooled from puromycin-resistant colonies produced 5-mm tumours within 18 days (data not shown).

Our results demonstrate that mutations of BRAF found in human cancers activate the kinase. Phosphorylation of residues within the activation segment regulates the activity of many kinases including BRAF¹⁰. In BRAF, T598 and S601 within the activation segment both require phosphorylation to achieve maximal kinase activity. These phosphorylations are effected after recruitment of BRAF to the membrane by activated RAS, and replacement of T598 and S601 by acidic amino acid residues results in RAS-independent activation of BRAF¹⁷. The V599 mutations found in human cancers probably mimic phosphorylation in the activation segment as they insert a negatively charged residue adjacent to a site of regulatory phosphorylation at S598. Notably, although many studies in experimental systems show that regulatory phosphorylations can be mimicked by substitution of an acidic amino acid, such mutations are probably rare in human disease, because an acidic amino acid substitution cannot be generated from a threonine or serine residue by a single base change. Our data now suggest that amino acids

other than threonine and serine can be mutated to acidic residues in human disease to mimic phosphorylation and hence activate kinases. Although the mode of action of V599 mutations may be explicable on the basis of mimicking regulatory phosphorylation, the mechanism of action of the other activation loop mutations is not clear and awaits structural studies. Moreover, the presence in one cancer cell line of an unusual double nucleotide substitution converting V599 to aspartic acid suggests as well that mutations at V599 may be particularly active biologically. Indeed, K600 could be changed by a single base substitution to glutamic acid, yet mutations at this residue have not been found in human cancers.

Our data also demonstrate that mutations of the glycine residues in the GXGXXG motif of the ATP-binding domain can activate kinases. It will be interesting to determine whether mutation of these conserved glycine residues activates other protein kinases and nucleotide-binding proteins. If this proves to be the case then it will provide a useful way to generate constitutively activated kinases experimentally.

The coding exons and intron–exon junctions of HRAS, KRAS and NRAS genes were screened for mutations through the entire panel of 545 cell lines. Seventy-one (13%) had mutations of RAS genes: 14 of 40 (35%) colorectal cancers, 23 of 131 (18%) lung cancers (22 NSCLC, 1 SCLC), 3 of 3 (100%) pancreatic cancers, 3 of 34 (9%) melanomas, 4 of 26 (15%) ovarian cancers, 3 of 27 (11%) neuroblastomas, 2 of 10 (20%) bladder cancers, 9 of 53 (17%) leukaemias/lymphomas and 10 of 221 others. Three of 43 cancer cell lines with BRAF mutations also had RAS gene mutations: BE, a colorectal cancer cell line (KRAS2 G13D; BRAF G463V); Hx62/26, an ovarian cancer cell line (KRAS2 G13D; BRAF G463E); and the NCI-H2087 NSCLC line (NRAS Q61K; BRAF L596V). Mutation screening of the 22 primary cancers with BRAF mutations for which adequate material was available revealed a colorectal cancer with coincident KRAS2 (G12V) and BRAF (F594L) mutations. The four cancers that had coincident RAS and BRAF mutations (3 cell lines, 1 primary tumour) were all from the set of 12 screened that had one of the less common BRAF mutations. Conversely, none of the 51 cancer samples with a V599 mutation screened through the RAS genes contained a RAS mutation (see Supplementary Information). This suggests that the common V599E mutation is biologically distinct from the other BRAF mutations.

Studies using injection of the RAS-neutralizing Y13-259 monoclonal antibody have previously shown that most normal cells examined and some tumour cells require RAS function for pro-

Table 3 Tumour cell lines containing the V599E BRAF mutant do not require RAS for proliferation

Cell line	Tissue	RAS mutation	BRAF mutation	Inhibition of S phase by Y13-259 (%)	Inhibition of S phase by U0126 (%)
WM-266-4	Melanoma	WT	V599D	10	99
SK-MEL-28	Melanoma	WT	V599E	4	98
A2058	Melanoma	WT	V599E	0	68
Malme	Melanoma	WT	V599E	0	ND
Colo741	Colorectal	WT	V599E	0	76
LS411N	Colorectal	WT	V599E	0	35
HT29	Colorectal	WT	V599E	15	ND
Colo205	Colorectal	WT	V599E	3	ND
Mawi	Colorectal	WT	V599E	5	8
NCI-H1666	NSCLC	WT	G465V	89	ND
BE	Colorectal	G13D KRAS2	G463V	97	ND
NCI-H2087	NSCLC	Q61K NRAS	L596V	77	56
Lim1899	Colorectal	G12A KRAS2	WT	74	ND
LS174T	Colorectal	G12D KRAS2	WT	84	ND
JW2	Colorectal	G12D KRAS2	WT	79	ND
SW620	Colorectal	G12V KRAS2	WT	92	92
DLD1	Colorectal	G13D KRAS2	WT	4	ND
HCT-116	Colorectal	G13D KRAS2	WT	95	75
SK-MEL-2	Melanoma	Q61R NRAS	WT	7	62
HMVI	Melanoma	Q61K NRAS	WT	100	86
CHL	Melanoma	WT	WT	96	51
SK-MEL-31	Melanoma	WT	WT	22	98

ND, not done; WT, wild type; NSCLC, non-small-cell lung cancers.

liferation in culture^{19,20}. Microinjection of Y13-259 into eight cell lines with the V599E mutation and the cell line with V599D did not block their proliferation in culture (Table 3). However, three cell lines tested with less common *BRAF* mutations (two of which had *RAS* mutations and one of which did not, Table 3) were inhibited by Y13-259 microinjection. These data suggest that *BRAF* V599 mutations uncouple cells from their proliferation requirement of *RAS* (although our *in vitro* data indicate that *BRAF* V599 mutants can be even further activated by mutant *RAS*), whereas other *BRAF* mutants remain dependent on *RAS* function. One interpretation of these results is that the less common *BRAF* mutants still require interaction with *RAS* to become phosphorylated and activated, whereas V599 mutants overcome the need for a *RAS*-dependent step by mimicking phosphorylation. The coincidence of *RAS* and *BRAF* mutants in the same cancer cell is, to our knowledge, the first report of tandem-activating mutations in more than one component of this signalling pathway. The observation prompts the speculation that mutant *RAS* signalling may be modulated by mutations at other locations in the pathway.

RAF proteins phosphorylate MEK1/2, which in turn phosphorylate ERK1/2. To evaluate whether activating mutations of *BRAF* signal through MEK and ERK, we treated cells with the MEK1/2 inhibitor U0126 (ref. 21). This compound inhibits DNA synthesis in a wide variety of cell types²¹. Of those cancer cell lines in which treatment with U0126 blocked ERK1/2 phosphorylation by at least 80%, 6 of 6 with either V599E or V599D mutations showed strong inhibition of DNA synthesis (Table 3). These results are therefore consistent with the hypothesis that the activated versions of *BRAF* signal, at least in part, through the classical MAPK cascade to promote proliferation.

Examination of the exons and splice junctions of *RAF1* in the set of 545 cancer cell lines did not reveal evidence of frequent mutations in melanoma or any other cancer. *BRAF* may be preferred as a mutational target because it has a higher basal kinase activity than *RAF1* (ref. 15). Indeed, when we introduced the *BRAF* mutations found in human cancers into the cognate positions in *RAF1*, they had at least 10-fold lower activity in kinase and transformation assays (data not shown). This may be due to the fact that *RAF1* requires phosphorylation on serine 338 and tyrosine 341 for activation of the kinase¹⁵. By contrast, in *BRAF*, the equivalent serine (S445) is constitutively phosphorylated and the position equivalent to the tyrosine is substituted by an aspartic acid residue (D448), which acts as a phosphomimetic. Thus *BRAF* may require fewer post-translational modifications than *RAF1* to achieve maximal kinase activity and hence is more susceptible to oncogenic activation.

We have identified *BRAF* as an oncogene in human cancer. The pattern and activity of mutations observed is likely to yield new insights into kinase function. There has recently been marked success reported of an inhibitor (STI571) of the BCR-ABL kinase in the treatment of chronic myeloid leukaemia²². The high frequency of *BRAF* mutations in melanoma and the relative lack of effective therapies for advanced stages of this disease suggest that inhibition of *BRAF* activity may be an important new strategy in the treatment of metastatic melanoma. The identification of *BRAF* as a commonly mutated target in human cancer at such an early stage of our genome-wide screen suggests that systematic searches through cancer cell genomes for somatic mutations ultimately will provide a much more complete picture of the number and patterns of mutations underlying human oncogenesis. □

Methods

Tissue samples

Normal and neoplastic tissue samples were obtained from publicly available banks in the case of cell lines, and from individual investigators using appropriate local Institutional Review Board approved protocols/ethical approval procedures for tissue collection in the case of primary tumour material. Cord blood controls were obtained from the North Cumbria Community Genetics Project (NCCGP). All 545 cancer cell lines and the 15

normal lymphoblastoid lines were genotyped using 400 polymorphic microsatellites (ABI linkage MD-10 panel) to confirm both matching of normal and tumour cell lines as well as non-duplication of cell lines.

Mutation screening

Screening for mutations was performed using a capillary-based modified heteroduplex method optimized to run on an ABI PRISM 3100 Genetic Analyser²³ (H.D., manuscript in preparation). PCR primers were designed to amplify the exon plus at least 50 bp of flanking intronic sequence (see Supplementary Information for primer sequences). A total of 12 ng genomic DNA from the test sample was mixed with 3 ng control genomic DNA and amplified using standard PCR conditions in which one of the primers was labelled with either FAM, NED or VIC dye. The resulting samples were then analysed on an ABI PRISM 3100 Genetic Analyser under semi-denaturing conditions using optimized separation medium and run conditions. The resulting traces were analysed using proprietary software to identify samples that produced a shift in peak migration relative to either the matched normal control from the same individual or a standard normal control, indicating the presence of a putative sequence variation. Samples that produced a heteroduplex shift were directly sequenced on both strands using BigDye terminator Cycle Sequencing Ready Reaction Kit (Applied Biosystems) according to the manufacturer's protocol, and analysed on an ABI PRISM 3100 Genetic Analyser.

RAS-neutralizing antibody microinjection and U0126 treatment

Tumour cells were seeded at a density such that they were approximately 50% confluent on the day of injection. Cells were microinjected with Y13-259 or control immunoglobulin- γ as described previously²⁴. Twenty hours after microinjection cells were labelled with 5-bromodeoxyuridine (BrdU) for 24 h, fixed and stained with antibody against BrdU. We scored at least 100 microinjected cells for each experiment. To determine the effects of inhibiting ERK1/2 activation on proliferation of tumour cells in culture, cells were seeded on day 1 then the following day U0126 was added in DMSO to 10 μ M; after 20 h BrdU was added for a further 24 h before fixation and staining with antibody against BrdU. Lysates were also made from cells treated in parallel to study the level of inhibition of ERK1/2 phosphorylation by western blotting with an antibody specific for the di-phosphorylated activated forms of ERK1/2 (clone MAP-YT; Sigma).

Transformation assays

1.3×10^5 cells of clone D4 of NIH 3T3 cells were transfected with 15–450 ng *BRAF* expression plasmids in pEFm6 (ref. 17) together with 50 ng pBabe Puro and sufficient empty pEFm6 to give a total of 700 ng DNA, using Lipofectamine (Invitrogen). After 24–26 h, cells were trypsinized and divided between two 10-cm tissue culture dishes containing DMEM plus 5% donor calf serum (focus assay) and two 10-cm dishes containing 2.5 μ g ml⁻¹ puromycin in DMEM plus 5% donor calf serum (transformed colony assay). The medium of the focus assays was changed every four days whereas the medium of the transformed colony assay was changed after seven days. Plates were scored, by experimenters 'blind' for transformation, after 12–18 days. 10^6 cells from pooled colonies of cells from the puromycin-selected dishes were injected subcutaneously into male nude mice aged over six weeks. Mice were observed twice weekly and killed when tumours reached a size in excess of 5 $\times$ 5 mm.

BRAF activity assays

Myc-epitope-tagged versions of *BRAF* were transiently expressed in COS cells using the reagent lipofectamine (Gibco/BRL) according to the manufacturer's instructions. Cell extracts were prepared and the Myc-epitope-tagged protein kinase activity was determined as described previously^{15,16}. Western blotting for ERK and pERK was performed as described^{15,16}.

Received 2 February; accepted 16 May 2002; doi:10.1038/nature00766.

Published online 9 June 2002.

1. Peyssonnaud, C. & Eychène, A. The Raf/MEK/ERK pathway: new concepts of activation. *Biol. Cell* **93**, 53–62 (2001).
2. Avruch, J. A. *et al.* Ras activation of the Raf kinase: tyrosine kinase recruitment of the MAP kinase cascade. *Recent Prog. Horm. Res.* **56**, 127–155 (2001).
3. Kolch, W. Meaningful relationships: the regulation of the Ras/Raf/MEK/ERK pathway by protein interactions. *Biochem. J.* **351**, 289–305 (2000).
4. Vogelstein, B. *et al.* Genetic alterations during colorectal-tumour development. *N. Engl. J. Med.* **319**, 525–532 (1988).
5. van't Veer, L. J. *et al.* N-ras mutations in human cutaneous melanoma from sun-exposed body sites. *Mol. Cell Biol.* **9**, 3114–3116 (1989).
6. Caduff, R. E., Svoboda-Newman, S. M., Ferguson, A. W., Johnston, C. M. & Frank, T. S. Comparison of mutations of Ki-RAS and p53 immunoreactivity in borderline and malignant epithelial ovarian tumours. *Am. J. Surg. Pathol.* **23**, 323–328 (1999).
7. Daya-Grosjean, L., Dumaz, N. & Sarasin, A. The specificity of p53 mutation spectra in sunlight induced human cancers. *J. Photochem. Photobiol. B* **28**, 115–124 (1995).
8. Halaban, R. The regulation of normal melanocyte proliferation. *Pigment Cell Res.* **13**, 4–14 (2000).
9. Busca, R. *et al.* Ras mediates the cAMP-dependent activation of extracellular signal-regulated kinases (ERKs) in melanocytes. *EMBO J.* **19**, 2900–2910 (2000).
10. Johnson, L. N., Lowe, E. D., Noble, M. E. & Owen, D. J. The Eleventh Datta Lecture. The structural basis for substrate recognition and control by protein kinases. *FEBS Lett.* **430**, 1–11 (1998).
11. Hemmer, W., McGlone, M., Tsigelny, I. & Taylor, S. S. Role of the glycine triad in the ATP-binding site of cAMP-dependent protein kinase. *J. Biol. Chem.* **272**, 16946–16954 (1997).
12. Grant, B. D., Hemmer, W., Tsigelny, I., Adams, J. A. & Taylor, S. S. Kinetic analyses of mutations in the glycine-rich loop of cAMP-dependent protein kinase. *Biochemistry* **37**, 7708–7715 (1998).
13. Odawara, M. *et al.* Human diabetes associated with a mutation in the tyrosine kinase domain of the insulin receptor. *Science* **245**, 66–68 (1989).

14. Cooke, M. P. & Perlmutter, R. M. Expression of a novel form of the *fyn* proto-oncogene in hematopoietic cells. *New Biol.* **1**, 66–74 (1989).
15. Mason, C. S. *et al.* Serine and tyrosine phosphorylations cooperate in Raf-1, but not B-Raf activation. *EMBO J.* **18**, 2137–2148 (1999).
16. Marais, R., Light, Y., Paterson, H. E., Mason, C. S. & Marshall, C. J. Differential regulation of Raf-1, A-Raf, and B-Raf by oncogenic ras and tyrosine kinases. *J. Biol. Chem.* **272**, 4378–4383 (1997).
17. Zhang, B. H. & Guan, K. L. Activation of B-Raf kinase requires phosphorylation of the conserved residues Thr598 and Ser601. *EMBO J.* **19**, 5429–5439 (2000).
18. Vojtek, A. B. & Der, C. J. Increasing complexity of the Ras signalling pathway. *J. Biol. Chem.* **273**, 19925–19928 (1998).
19. Mulcahy, L. S., Smith, M. R. & Stacey, D. W. Requirement for ras proto-oncogene function during serum-stimulated growth of NIH 3T3 cells. *Nature* **313**, 241–243 (1985).
20. Stacey, D. W., DeGudicibus, S. R. & Smith, M. R. Cellular ras activity and tumour cell proliferation. *Exp. Cell Res.* **171**, 232–242 (1987).
21. Favata, M. F. *et al.* Identification of a novel inhibitor of mitogen-activated protein kinase kinase. *J. Biol. Chem.* **273**, 18623–18632 (1998).
22. Druker, B. J. *et al.* Efficacy and safety of a specific inhibitor of the BCR-ABL tyrosine kinase in chronic myeloid leukaemia. *N. Engl. J. Med.* **344**, 1031–1037 (2001).
23. Rozycka, M., Collin, N., Stratton, M. R. & Wooster, R. Rapid detection of DNA sequence variants by conformation sensitive capillary electrophoresis. *Genomics* **70**, 34–40 (2000).
24. Mittnacht, S., Paterson, H., Olson, M. F. & Marshall, C. J. Ras signalling is required for inactivation of the tumour suppressor pRb cell-cycle control protein. *Curr. Biol.* **7**, 219–221 (1997).

Supplementary Information accompanies the paper on Nature's website (<http://www.nature.com/nature>).

Acknowledgements

We would like to thank all the patients who donated samples for these studies, the UK Children's Cancer Study Group for provision of paediatric primary tumour samples, the NCCGP for provision of cord blood control DNA samples, and W. Haynes for assistance with preparation of the manuscript. We would also like to acknowledge the Wellcome Trust, Institute of Cancer Research and Cancer Research UK for support. C.J.M. is a Gibb life fellow of the Cancer Research UK. G.P. and A.C. are funded in part by Regione Autonoma della Sardegna. B.A.G. is supported by Breakthrough Breast Cancer.

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to R.W. (e-mail: rw1@sanger.ac.uk).

VEGF regulates haematopoietic stem cell survival by an internal autocrine loop mechanism

Hans-Peter Gerber*, Ajay K. Malik*, Gregg P. Solar†, Daniel Sherman*, Xiao Huan Liang*, Gloria Meng‡, Kyu Hong‡, James C. Marsters§ & Napoleone Ferrara*

Departments of * Molecular Oncology, † Protein Chemistry, ‡ Assay Development and § Organic Chemistry, Genentech, Inc., 1 DNA Way, South San Francisco, California 94080, USA

Vascular endothelial growth factor (VEGF) is a principal regulator of blood vessel formation and haematopoiesis^{1,2}, but the mechanisms by which VEGF differentially regulates these processes have been elusive. Here we describe a regulatory loop by which VEGF controls survival of haematopoietic stem cells (HSCs). We observed a reduction in survival, colony formation and *in vivo* repopulation rates of HSCs after ablation of the VEGF gene in mice. Intracellularly acting small-molecule inhibitors of VEGF receptor (VEGFR) tyrosine kinase dramatically reduced colony formation of HSCs, thus mimicking deletion of the VEGF gene. However, blocking VEGF by administering a soluble VEGFR-1, which acts extracellularly, induced only minor effects. These findings support the involvement in HSC survival of a VEGF-dependent internal autocrine loop mechanism (that is, the mechanism is resistant to inhibitors that fail to penetrate the

intracellular compartment). Not only ligands selective for VEGF and VEGFR-2 but also VEGFR-1 agonists rescued survival and repopulation of VEGF-deficient HSCs, revealing a function for VEGFR-1 signalling during haematopoiesis.

Differentiation, maintenance and expansion of HSCs is part of a highly orchestrated process involving multiple growth factors, cytokines and chemokines, which act in complex circuits of paracrine and autocrine regulation. Stromal cells such as fibroblasts, macrophages, T lymphocytes and endothelial cells residing in the bone marrow secrete a range of cytokines regulating the maintenance of HSCs as well as their rapid expansion during pathologic conditions. The apparent physical association of HSCs with stromal cells at sites of HSC maintenance and differentiation suggested paracrine and juxtacrine (membrane-anchored) mechanisms. However, various growth factors and cytokines including interleukin (IL)-3, granulocyte-macrophage colony-stimulating factor (GM-CSF), IL-6 and VEGF are co-expressed with their respective receptors on normal, early and differentiated haematopoietic cells, suggesting that autocrine mechanisms are involved in the regulation of haematopoiesis (for review see ref. 3).

VEGF is expressed in bone marrow, and VEGF levels in HSCs increase in response to cytokine stimulation⁴. VEGFR-2 is present

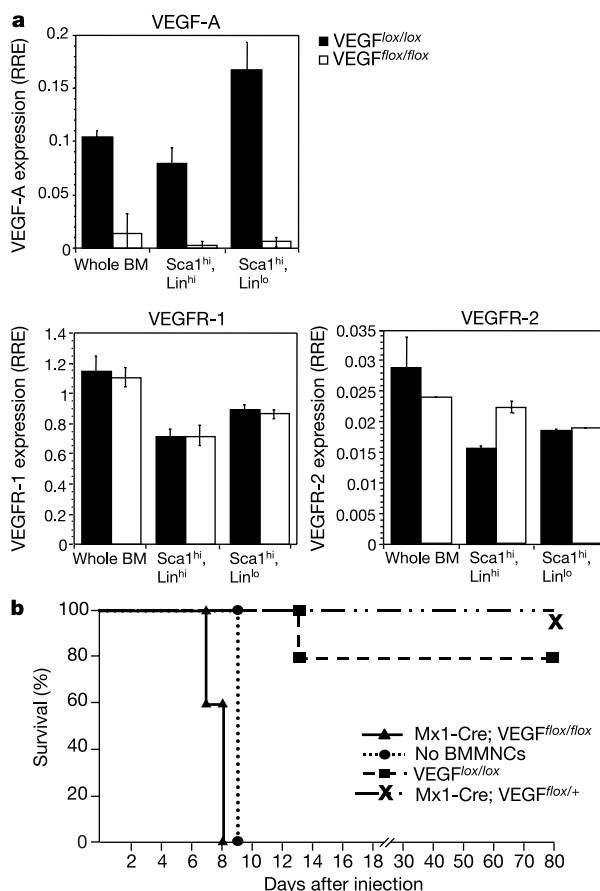


Figure 1 Real-time RT-PCR analysis of BMMNCs and HSCs and competitive repopulation frequencies of VEGF-deficient HSCs. **a**, VEGF^{lox/lox} cells are derived from VEGF^{lox/lox} cells following gene ablation. Total RNA was isolated from 1×10^6 VEGF^{lox/lox} and VEGF^{lox/lox+} cells BMMNCs or HPCs (Sca⁺) that were lineage committed (Lin^{hi}) or not (Lin^{lo}) 4 d after addition of IFN- α . Relative RNA units (RRE) for VEGFR-1 (Fit-1), VEGFR-2 (KDR/Flk-1) and VEGF-A were normalized to glyceraldehyde-3-phosphate dehydrogenase (GAPDH) levels and calculated from standard curves²⁸. Data shown are means $\pm$ s.d. from three RNA preparations. BM, bone marrow. **b**, Survival experiment of lethally irradiated C57BL/6 mice (five mice per group) repopulated with wild-type (VEGF^{lox/lox+}), heterozygous (Mx1-Cre; VEGF^{lox/lox+}) and VEGF-deficient (Mx1-Cre; VEGF^{lox/lox}) BMMNCs.

Table 1 Competitive repopulation assay with HPCs and BMMNCs

Genotype	4 weeks		8 weeks		28 weeks	
	HPCs (%)	BMMNCs (%)	HPCs (%)	BMMNCs (%)	HPCs (%)	BMMNCs (%)
VEGF ^{flx/flx}	0.2 ± 0.01	1.1 ± 1.0	1.2 ± 0.9	0	1.3 ± 0.8	0
VEGF ^{lox/lox}	72.9 ± 6.9	15.8 ± 3.6	36.4 ± 6.2	23.1 ± 5.1	24.9 ± 4.3	20.1 ± 4.9

Repopulation frequencies of cell isolates from Mx1-Cre⁺ VEGF^{flx/flx} or Mx1-Cre⁺ VEGF^{lox/lox} littermates. Long-term repopulation analysis of CD45.2⁺ bone marrow progenitors Sca1⁺ Lin^{lo} HPCs or BMMNCs was conducted in C57BL/6 CD45.1⁺ host mice. Data shown are the percentage of cells double positive for CD45.2 and lineage markers. BMMNCs have lower repopulation potential than the Sca1⁺ Kit⁺ Lin^{lo} HPC fractions because of their comparatively lower content of stem cells. Data are the means ± s.e.m. of five mice.

on a subset of multipotent haematopoietic stem cells⁵, and long-term *in vitro* cell culture assays and *in vivo* studies indicate that pluripotent stem cell activity correlated with VEGFR-2 expression⁶. A role for VEGF and VEGFR-2 in haematopoiesis is further supported by conventional gene knockout experiments, which resulted in early embryonic lethality owing to impaired haematopoiesis and angiogenesis^{1,2,5}. However, such prenatal lethality precluded the analysis of either process during development and hampered the dissection of the mechanisms controlling haematopoiesis and angiogenesis in adults.

Conditional Cre-loxP-based gene knockout technology, combined with haematopoietic repopulation experiments, enabled us to separate the relative contributions of autocrine and paracrine VEGF effector mechanisms during haematopoietic engraftment, lineage differentiation and cell survival in adult mice, independently of VEGF's angiogenic functions⁷. We have previously described the generation of transgenic mice carrying loxP sites flanking exon 3 of the VEGF gene and a transgene for Cre recombinase, the expression of which is controlled by an interferon (IFN)-α inducible promoter (Mx1-Cre)⁸. To investigate the role of VEGF in HSCs, we isolated bone marrow mononuclear cells (BMMNCs), haematopoietic progenitor cells (HPCs) (Sca1⁺ Kit⁺ Lin^{lo}) and lineage-committed progenitor cells (Sca1⁺ Lin^{lo}) from transgenic mice. We induced VEGF gene ablation in these cells *in vitro* by transient exposure to IFN-α (400 U ml⁻¹) for 3–4 days in the presence of exogenous VEGF (50 ng ml⁻¹). Real-time polymerase chain reaction with reverse transcription (RT-PCR) (Taqman) demonstrated a greater than 95% reduction in VEGF expression in the absence of alterations in the expression levels of VEGFR-1 or -2 (Fig. 1a).

Next, we performed bone marrow transplantation experiments in which VEGF-deficient or control bone marrow cells were introduced into lethally irradiated wild-type recipient mice by injection into the tail vein. Administration of VEGF-deficient BMMNCs (VEGF^{flx/flx}) failed to restore the haematopoietic compartment and all mice died 7–8 days after reconstruction (Fig. 1b). Mice receiving control BMMNCs heterozygous for VEGF (VEGF^{flx/+}) or wild-type cells (VEGF^{+/+}) mostly survived. To circumvent the lethality caused by VEGF-deficient BMMNCs, we complemented

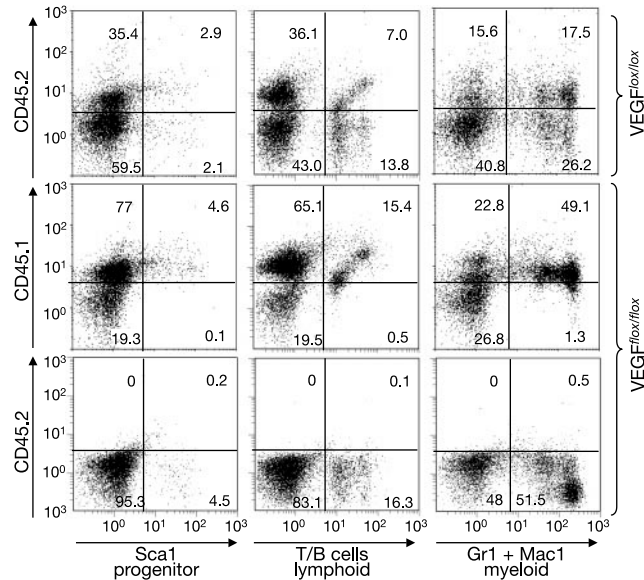


Figure 2 Comparison between the competitive repopulation of CD45.2-positive Sca1⁺ Kit⁺ Lin^{lo} bone marrow cells deleted in both VEGF alleles (VEGF^{flx/flx}) or control (VEGF^{lox/lox}) cells 8 weeks after transplantation into lethally irradiated CD45.1-positive recipient mice. Data shown are representative FACS chromatograms of BMMNCs from one of five mice per group. To assess the repopulation frequencies of the donor cells, cell isolates were double stained with antibodies recognizing CD45.2 or CD45.1 and B220, CD4, CD8 (lymphoid lineages), Gr-1/Mac-1 (myeloid lineages) and Sca1 (progenitors), respectively⁹. There was no contribution of VEGF^{flx/flx} CD45.2 cells to any of the lineages analysed.

donor cells with excess wild-type host cells before injection⁹. Donor or host cells are congenic at the CD45 locus, so their relative repopulation frequencies can be assessed by using antibodies that recognize donor (CD45.2) and host (CD45.1) markers, respectively. When tested in such a competitive repopulation strategy, all VEGF-deficient (VEGF^{flx/flx}, Sca1⁺ Kit⁺ Lin^{lo}) HPCs as well as BMMNCs failed to repopulate lethally irradiated host mice (Tables 1 and 2). Peripheral blood cell samples and bone marrow cell isolates were analysed for co-expression of CD45.2 and lineage markers for the lymphoid (CD4, CD8, B220), myeloid (Gr1/Mac1) and progenitor (Sca1) compartments; these were entirely absent (Fig. 2). Control cells repopulated at normal frequencies and we did not detect any differences between cells that had undeleted VEGF alleles (VEGF^{lox/lox}) and cells heterozygous for VEGF (Table 2, VEGF^{flx/+}).

To test the possibility that impaired homing of BMMNCs to the recipient bone marrow may be responsible for the failure of VEGF-deficient cells to repopulate the host, we analysed the bone marrow of mice 1 day after administration of 20 × 10⁶ VEGF-deficient or control cells. Owing to the massive increase in autofluorescence induced by lethal irradiation, these experiments were conducted in

Table 2 Colony formation and repopulation after retroviral transduction

	No virus	MSCV, IRES, GFP, empty virus	MSCV, IRES, GFP, VEGF 165	MSCV, IRES, GFP, PIGF	No virus + Fit-IgG (25 mg kg ⁻¹)
<i>In vitro</i> colony formation (total colony numbers)					
VEGF ^{flx/flx} CRE ⁺	12 ± 3	14 ± 2	112 ± 14	72 ± 6	ND
VEGF ^{lox/lox} CRE ⁻	121 ± 8	108 ± 4	142 ± 6*	123 ± 10	ND
VEGF ^{+/+} CRE ⁻	101 ± 6	ND	ND	ND	81 ± 13
<i>In vivo</i> repopulation frequency (%)					
VEGF ^{flx/flx} CRE ⁺	0.3 ± 0.07	0.4 ± 0.09	77.4 ± 5.5	69.8 ± 1.0	ND
VEGF ^{lox/lox} CRE ⁻	72.1 ± 0.7	68.9 ± 0.6	81.8 ± 2.6†	66.4 ± 1.4	81.2 ± 1.7
VEGF ^{flx/+} CRE ⁺	72.6 ± 1.1	ND	ND	ND	ND

For *in vitro* colony formation, Sca1⁺ Kit⁺ Lin^{lo} HPCs were serially transduced with retroviral constructs for 6 d and then cultured for 3 d in the presence of IFN-α (400 U ml⁻¹) and VEGF (30 ng ml⁻¹). There were no significant differences in the colony formation frequencies between cells exposed to virus or IFN-α compared with control cells (112 ± 7 colonies). Data shown are means ± s.e.m. of total colony numbers of quadruplicate plates from one representative of two independent experiments. For *in vivo* repopulation assays, data shown are means ± s.e.m. of the percentage of cells double positive for CD45.2 and all lineage markers from five animals per group 4 weeks after repopulation. ND, not determined.

*P < 0.05 and †P < 0.005 compared with the empty virus group.

non-irradiated mice. Fluorescence-activated cell sorting (FACS) for CD45.2⁺ donors among the CD45.1⁺ host bone marrow cells revealed no significant differences between VEGF-deficient (3.8 ± 0.9%, *n* = 4), heterozygous (3.7 ± 1.3%, *n* = 4) and wild-type BMMNCs (5.2 ± 1%, *n* = 4) 1 day after injection. Thus, the microenvironment consisting of wild-type stromal cells is sufficient for homing of VEGF-deficient BMMNCs. However, it fails to rescue VEGF-deficient BMMNCs and HPCs or to promote survival and differentiation during haematopoietic repopulation, suggesting that a cell-autonomous mechanism underlies the function of VEGF during such events.

Another hallmark of HPC activity is their capacity to form colonies *in vitro*, which can be morphologically phenotyped under a light microscope as 'burst-forming unit-erythroid' (BFU-E), 'colony-forming unit-granulocyte/macrophage' (CFU-GM), 'colony-forming unit-granulocyte/erythroid/monocyte/megakaryocyte' (CFU-GEMM) or macrophage¹⁰. The ability of VEGF-deficient Sca⁺ Kit⁺ Lin^{lo} HPCs to form colonies *in vitro* was dramatically reduced. All progenitors were equally affected and there were no indications of phenotypic skewing (Fig. 3a). Most cells developed the characteristic morphological changes associated with apoptotic cell death, including membrane blebbing and nuclear condensation (data not shown). Similarly, we found a decrease of about 80% in colony formation rates of BMMNCs and Sca⁺ Kit⁺ Lin^{lo} HPCs when VEGF gene ablation had been induced *in vivo* by systemic administration of poly(I:C), a compound that induces production of endogenous IFN-α. In these experiments, colony formation rates of control VEGF^{lox/lox} or heterozygous VEGF^{lox/+} cells were comparable to wild-type

Table 3 Effect of poly(I:C) treatment on *in vitro* colony formation

Cell type	PBS		Poly(I:C)	
	VEGF ^{lox/lox} Cre ⁺	VEGF ^{lox/lox} Cre ⁻	VEGF ^{lox/lox} Cre ⁺	VEGF ^{lox/lox} Cre ⁻
BMMNCs	78 ± 13	82 ± 9	7 ± 2	90 ± 15
HPCs	58 ± 9	65 ± 10	10 ± 4	60 ± 5

Sca⁺ Kit⁺ Lin^{lo} HPCs or BMMNCs were isolated from Mx1-Cre⁺ VEGF^{lox/lox} or Mx1-Cre⁻ VEGF^{lox/lox} control littermates 1 week after treatment with poly(I:C) or PBS control and tested for their colony formation potential. Cells were plated on methylcellulose media and counted and phenotyped after 10–14 d as described in Fig. 2. Data shown are the means ± s.d. of total colony number from duplicate plates derived from one representative of three independent experiments.

VEGF^{+/+} cells (Table 3). Thus, our findings argue against the possibility that the increase in apoptosis in VEGF-deficient HSCs may be related to the specific *in vitro* manipulations used to induce gene deletion, because the deficit occurred irrespective of the environment in which the VEGF deficiency had been created.

However, a soluble VEGFR-1-immunoglobulin-γ (IgG) chimaeric protein, mouse Flt1-3-IgG (ref. 8), failed to induce HSC apoptosis and mimic VEGF gene deletion, even when added at concentrations more than 500-fold higher than those required to inhibit the mitogenic activity of 5 ng ml⁻¹ VEGF (Table 4, Fig. 3d). These findings suggest that the VEGF responsible for the survival activity is not accessible to inhibitors that act in the extracellular compartment. In this context, it is noteworthy that treatment of adult mice with high doses of Flt1-3-IgG for 4 weeks failed to induce any alterations in peripheral blood cell counts (ref. 8 and data not shown). The lack of effect of Flt1-3-IgG in HSC survival is consistent with a 'private' or internal autocrine loop mechanism, as previously described for IL-3 (ref. 11) and GM-CSF (ref. 12). A

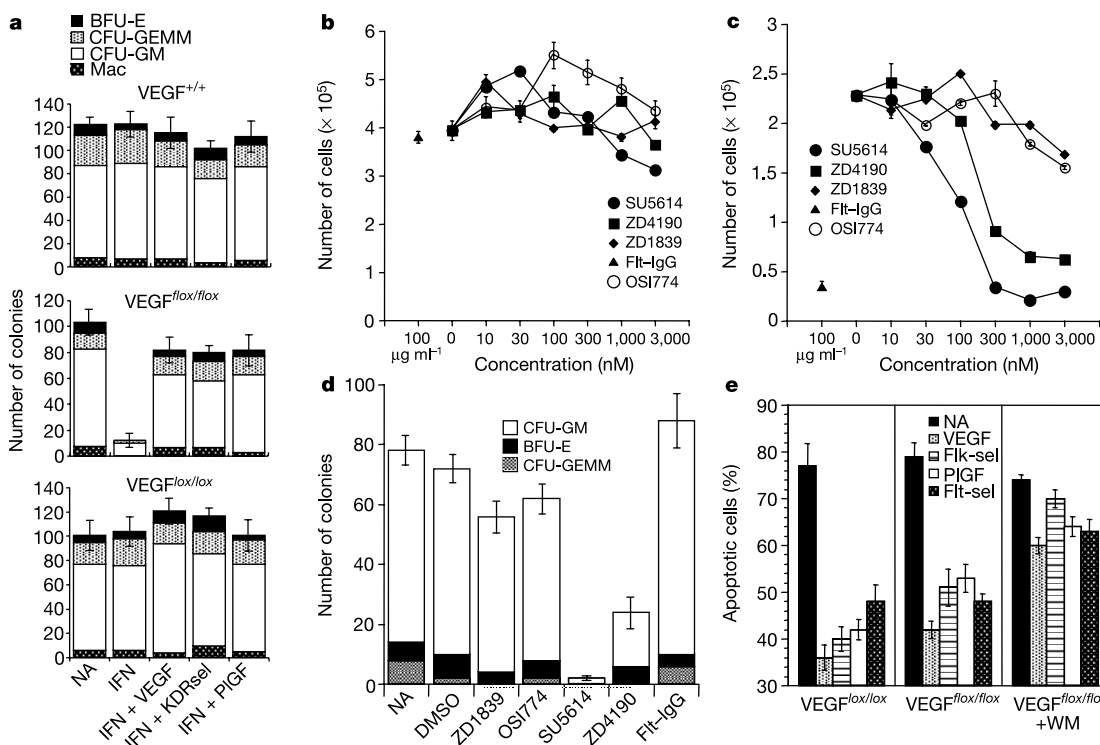


Figure 3 VEGF inactivation by Flt-IgG interferes with the paracrine but not autocrine mechanism and small-molecule inhibitors targeting VEGFR-1 and -2 block both functions. **a**, VEGF^{+/+}, VEGF^{lox/lox} or VEGF^{lox/+} (Sca⁺ Lin^{lo}) bone marrow cells were cultured in the presence of IFN-α (400 U ml⁻¹) as indicated. All colony formation data represent the means ± s.e.m. of quadruplicate plates from one representative of three independent experiments. **b**, Primary HUVECs were grown in complete growth medium including small-molecule inhibitors as indicated. Total cell numbers were assessed after 5 d of growth. **c**, HUVECs were cultured in 1% FCS with 50 ng ml⁻¹ of VEGF in the presence of

the indicated compounds and counted after 5 d of growth. **d**, Sca⁺ Kit⁺ Lin^{lo} HPCs were plated in methylcellulose with 3 μM of inhibitor compounds or 20 μg ml⁻¹ Flt-IgG and counted by phenotype after 12–14 d. NA, no addition. **e**, Sca⁺ Lin^{lo} HPCs were incubated in the presence of 50 ng ml⁻¹ of each of the indicated growth factors or wortmannin (WM, 100 nM). Data represent the percentage of annexin-V-positive, PI-negative cells from one representative of three independent experiments. Error bars are ± s.d. of triplicate samples.

Table 4 *In vitro* clonogenic assay

IFN- α	—	—	+	+	+	+
VEGF	—	—	—	—	+	+
Flt-IgG	—	+	—	+	—	+
VEGF ^{flax/lox} Cre ⁺	89 $\pm$ 8	72 $\pm$ 8	12 $\pm$ 6	12 $\pm$ 4	82 $\pm$ 11	18 $\pm$ 8
VEGF ^{lox/lox} Cre ⁺	101 $\pm$ 10	103 $\pm$ 7	98 $\pm$ 14	99 $\pm$ 12	ND	ND
VEGF ^{flax/lox} Cre ⁺	85 $\pm$ 12	ND	101 $\pm$ 16	ND	90 $\pm$ 12	ND

Freshly isolated Sca⁺ Lin^{lo} VEGF^{flax/lox}, VEGF^{lox/lox} or VEGF^{flax/lox} bone marrow cells were cultured in the presence of IFN- α (400 U ml⁻¹), human VEGF (50 ng ml⁻¹) and/or mouse Flt-IgG (20 μ g ml⁻¹) as indicated. For *in vitro* colony formation, BMCs were plated in methylcellulose and counted after 12–14 d. Data shown are the means $\pm$ s.e.m. colony numbers of quadruplicate plates derived from one representative of two independent experiments. Similar to the data shown in Fig. 2a, all progenitor populations analysed were equally affected and there was no indication of phenotypic skewing. ND, not determined.

‘private’ (as opposed to a ‘public’) autocrine loop is one that is refractory to inhibitors that fail to penetrate the intracellular compartment, such as antibodies. However, intracellularly acting inhibitors of receptor signalling are expected to block it. In agreement with this hypothesis, two structurally unrelated small-molecule blockers of both VEGFR-1 and VEGFR-2 tyrosine kinase activity (ZD4190 (ref. 13) and SU5416 (ref. 14)) substantially reduced *in vitro* rates of HSC colony formation, whereas two control inhibitors targeting epidermal growth factor (EGF) receptor family members (ZD1839 and OSI774) did not impair HSC colony formation significantly (Fig. 3d). Exposure of primary human umbilical vein endothelial cells (HUVECs) to either compound did not induce any significant cellular toxicity (Fig. 3b), but caused the expected inhibition of proliferation in response to exogenous VEGF (Fig. 3c). These findings suggest that a VEGF-dependent autocrine survival loop does not generally operate in endothelial cells and may be restricted to HSCs.

It is noteworthy that inactivation of one single VEGF allele (VEGF^{flax/+}) did not alter repopulation frequencies significantly (Table 2), whereas it had a dramatic impact on embryonic development, causing early embryonic lethality¹. A potential explanation for this apparent discrepancy is that a purely autocrine survival mechanism is less dose dependent than the generation of complex paracrine gradients of VEGF concentration. Because these latter processes are required for blood vessel formation during organ development, our data are consistent with the model in which the mechanisms regulating angiogenesis and haematopoiesis are fundamentally different.

We next wished to determine the differential role of the two VEGF receptors in mediating VEGF-dependent HSC survival. We tested VEGF or a VEGFR-2-selective mutant (KDR-sel), and Flt-sel and placenta growth factor (PlGF), both of which specifically interact with VEGFR-1 (refs 15, 16), in an *in vitro* clonogenic assay. Addition of wild-type VEGF or KDR-sel restored survival and differentiation in VEGF-deficient HSCs *in vitro* (Table 4). This is in agreement with reports that provided evidence for the requisite role of VEGFR-2 during haematopoietic differentiation *in vivo*⁵. In contrast, the role of VEGFR-1 so far has been rather enigmatic. This molecule signals poorly in endothelial cells and has even been reported to have inhibitory effects by acting as a ‘decoy’ receptor or even directly inhibiting VEGFR-2 activities¹⁷, although other studies suggest that VEGFR-1 may mediate recruitment of monocytes¹⁸ and endothelial cell progenitors to the tumour vasculature¹⁹. However, to our surprise, PlGF and Flt-sel (data not shown) rescued VEGF-deficient HSCs with similar potency as wild-type VEGF or KDR-sel, indicating that both VEGF receptors are able to mediate this effect (Fig. 3a, e). Furthermore, we performed *in vitro* clonogenic and *in vivo* repopulation experiments with purified HPCs transduced with bicistronic retroviral vectors coding for either VEGF or PlGF and GFP. Enforced expression of VEGF or PlGF not only restored the ability of VEGF-deficient HSCs to form colonies *in vitro*, but also increased the colony formation rates of wild-type HSCs. *In vivo*, enforced expression of either ligand rescued the repopulation potential of VEGF-deficient HSCs and

increased the repopulation frequencies of wild-type HSCs (Table 2). To extend these observations, we incubated serum-starved HPCs with receptor-selective VEGF ligands and determined the levels of apoptosis by FACS analysis for annexin-V binding. Sca⁺ Kit⁺ Lin^{lo} HPCs rapidly underwent apoptotic cell death starting 24 h after onset of serum starvation (Fig. 3e). Such decrease in cell survival was potentially inhibited by receptor-selective ligands for both VEGF receptors in a manner dependent on phosphoinositide 3-kinase (PI3K), as demonstrated by the inhibitory effect of the PI3K inhibitor wortmannin. This is in contrast to previous studies in endothelial cells, where most of the VEGF activity is mediated by VEGFR-2, and VEGFR-1-selective ligands failed to induce any appreciable survival effect²⁰. It will be of considerable interest to identify the differences between endothelial cells and HSCs in regard to VEGFR-1 signalling that account for such marked differences.

Numerous growth factors are known to be produced by normal human CD34⁺ bone marrow cells, and on the basis of receptor co-expression, autocrine loops in HSCs have been proposed for IL-1, c-Kit ligand, Flt3 ligand and erythropoietin (EPO)²¹. The discovery of an autocrine loop formed by VEGF and its receptors might have an impact on the therapeutic strategies used for the treatment of haematological malignancies, which often represent clonal derivatives of HPCs. Increased expression levels of VEGF and VEGFR-1 and -2 have been found in many human haematopoietic tumour cell lines and tissues representing several lineages and diseases, including leukaemia, lymphoma and multiple myeloma^{22,23}, and there is evidence for a role for VEGFR-2 in regulating invasiveness of xenotransplanted human leukaemia cell lines in immunocompromised mice²⁴. However, interference with this loop may, at least in part, be responsible for certain unexpected effects of small-molecule VEGFR kinase inhibitors^{25,26}. For example, a significant reduction in haematocrit in adult rodents treated with these inhibitors has been reported²⁵, whereas a similar effect had not been observed in studies with anti-VEGF antibodies or soluble receptors^{8,27}. Our findings elucidate a mechanism by which VEGF regulates HSC survival. The relevance of these findings to the design of anti-angiogenic strategies awaits further study. □

Methods

VEGF gene ablation *in vivo* and *in vitro*

BMMNCs and HSCs were isolated as described previously⁹ and grown in IMDM with 10% fetal calf serum (FCS) in the presence of 400 U ml⁻¹ recombinant mouse (rm) IFN- α (10,000 U, Chemicon), 50 ng ml⁻¹ Kit ligand (Chemicon), rm thrombopoietin (TPO) (Genentech), 50 ng ml⁻¹ rm stem-cell factor (SCF) (R&D Systems), 50 ng ml⁻¹ Flt3 ligand (Immunex) and recombinant human (rh) VEGF (50 ng ml⁻¹, Genentech) for 3–4 d. Knockout efficiencies were determined by PCR analysis of genomic DNA or real-time RT-PCR with total RNA isolated from *in vitro* cultures as previously described⁸. Efficiency of VEGF gene ablation was monitored at the level of DNA recombination by Southern blot analysis and was routinely found to be >95% (ref. 8 and data not shown). For Southern blot analysis, genomic DNA was isolated from BMMNCs or Lin⁺ BMMNCs and digested with *Pvu*II. A 1-kilobase probe was generated by PCR using genomic DNA as template with the primers mgVEGF175F (ACCAGCAGCCAAAGGAAAGCT) and mgVEGF1194R (TGCGCTCTGACTCTTTGC). For *in vivo* VEGF gene ablation experiments, mice were injected intraperitoneally three times, every third day with 300 μ g of poly(I:C) (Sigma). Before injection into lethally irradiated recipient mice, cells were analysed for viability by trypan blue dye exclusion or annexin-V binding by FACS analysis as previously described²⁰ (data not shown).

Endothelial cell proliferation assay

HUVECs and endothelial growth media were purchased from Clonetics. Cells were plated in complete media at a density of 30,000 cells per well in six-well dishes and, 24 h later, the medium was changed to 1% FCS complemented with the indicated compounds for 5 d before cell counting as described previously²⁰.

In vitro clonogenic assay

Methylcellulose cultures were initiated in 35-mm gridded plates with 4 $\times$ 10³ HPCs or 6.7 $\times$ 10⁵ BMMNCs per duplicate plate in complete methylcellulose containing 50 ng ml⁻¹ rmSCF, 10 ng ml⁻¹ rmlL-3, 10 ng ml⁻¹ rhIL-6, and 3 U ml⁻¹ rhEPO (Stem Cell Technologies). Colonies were scored and phenotyped on an inverted phase microscope 11–14 d after seeding as previously described⁹. Culture medium was supplemented with 50 ng ml⁻¹ of rhVEGF, 50 ng ml⁻¹ of a KDR-selective mutant VEGF form, 50 or 500 ng ml⁻¹ of recombinant human or mouse PlGF (R&D Systems), or 400 U ml⁻¹ of rmlIFN- α (Chemicon) where indicated. Flt1-3-IgG was purified as previously described²⁸.

At 30 ng ml⁻¹, this preparation was able to fully neutralize 5 ng ml⁻¹ VEGF. The endotoxin level was <1 endotoxic unit per mg. The KDR-sel mutant was VEGF D63S/G65M/L66R, and Flt-sel was VEGF I43A/I46A/Q79A/I83A, as described¹⁵. Receptor-selective binding of all ligands to the murine receptors was confirmed by enzyme-linked immunosorbent assay (ELISA) using murine VEGFR-1 and -2 protein (data not shown). ZD4190 was prepared by methods described in the literature²⁹ using Mitsunobu coupling of hydroxyethyl-1-triazole to C-7 of the anilinoquinazoline followed by precipitation of the HCl salt (purity >97% by HPLC, identity to published material assessed by ¹H nuclear magnetic resonance (NMR) and liquid chromatography/mass spectrometry (LCMS)). SU5416 was prepared by the literature method (purity >96% by HPLC, identity to published material assessed by ¹H NMR and LCMS). Given the fact that at least SU5416 was previously shown to inhibit c-Kit activation by SCF³⁰, in parallel experiments we omitted SCF from the colony formation assay and replaced it with mouse Flt3 ligand (50 ng ml⁻¹) and TPO (50 ng ml⁻¹). Under these conditions, SU5416 and ZD4190 induced respectively a 79.1% and 60.4% reduction in colony formation when added at 3 μM concentration, indicating that the decrease in colony formation by the small-molecule inhibitors can occur independently of their effects on c-Kit activity.

Retroviral engraftment and cell repopulation assays

Retroviral vectors coding for murine VEGF, murine PlGF or empty vector in tandem with an internal ribosome entry site (IRES) and green fluorescent protein (GFP) were transiently transfected into the packaging line Phoenix-ampho. Virus titres were determined to be 2–4 × 10⁶ plaque-forming units (PFU) per ml of cell suspension. Supernatants containing retrovirus were collected, sterile filtered (0.45 μm) and added to cells in presence of rmSCF 50 ng ml⁻¹, rmFlt3 ligand 50 ng ml⁻¹ (R&D Systems), rmTPO 50 ng ml⁻¹ (Genentech) and polybrene (8 μg, Sigma). We exposed 1 × 10⁵ Sca⁺ Kit⁺ Lin^{lo} HPCs or 1 × 10⁸ BMMNCs four times to fresh virus preparations, cytokines and polybrene during four consecutive days. GFP-positive cells were sorted by FACS and administered to lethally irradiated CD45.1⁺ adult male C57BL/6 recipient mice by tail vein injection. Whole-body irradiation (1,100 rad, 780 cGy min⁻¹) was administered as single dose from a ¹³⁷Cs source. Five animals per group were used. We used 1 × 10⁶ BMMNCs from CD45.1⁺ host mice as a source competitor, and mixed them with 1 × 10⁴ CD45.2⁺ Sca⁺ Kit⁺ Lin^{lo} donor HPCs as previously described⁹. For bone marrow engraftment studies, 1–2.5 × 10⁷ BMMNCs were administered into non-irradiated CD45.1⁺ recipient mice. One day after injection, BMMNCs and peripheral blood mononuclear cells (PBMCs) were isolated and analysed by FACS for CD45.2 expression. Peripheral blood samples were obtained from the retro-orbital sinus 4, 12 and 28 weeks after reconstitution. The percentage of CD45.2⁺ donor cells was determined by staining with biotin-conjugated CD45.2 monoclonal antibodies (A20.1.17) as described previously⁹. Data are shown as the percentage of donor (CD45.2) relative to recipient (CD45.1) bone marrow cells as determined from FACS data. To test for stromal cell contaminants, we analysed HPC preparations for the CD31/PECAM marker, which is indicative for endothelial cells. The absence of CD31/PECAM⁺ cells in Sca⁺ Lin^{lo} or c-Kit⁺ Sca⁺ Lin^{lo} isolates, and the notion that stromal cells from adults do not express c-Kit or Sca1, strongly argue against the presence of stromal cells in our HPC preparations (data not shown).

Flow cytometric analysis of bone marrow cells

Mice were euthanized and bone marrow cells were isolated as described before⁹. Briefly, bone marrow cells were isolated by flushing femurs with 2% FCS in PBS and a single-cell suspension was made. Remaining red blood cells were removed by lysis in 10 mM NH₄Cl. The mononuclear fraction (BMMNC) was then incubated with monoclonal antibodies binding to lineage-specific markers including rat anti-mouse CD4/L3T4, CD8/Ly2,3, B-220/CD45 (Caltag), GR1, Mac-1 and Ter-119 (Pharmingen). For purification of HPCs, BMMNCs were depleted of lineage-committed cells by incubation with the respective antibodies followed by magnetic bead depletion (Miltenyi Biotec) and subsequent processing on an automated magnetic cell-sorting machine. The unbound fraction was further sorted for Kit and/or Sca1 double-positive cells on a FACS machine (Epics Elite ESP). Annexin-V-fluorescein isothiocyanate (FITC) (BioVision Research) was used to determine apoptosis. For serum-starvation experiments, cells were kept in serum-free medium complemented with SCF (R&D Systems, 50 ng ml⁻¹), TPO (Genentech, 50 ng ml⁻¹) and Flt3 (R&D Systems, 50 ng ml⁻¹) for the indicated periods. Wortmannin (Sigma) was added 2 h before and during growth factor stimulation.

Total RNA isolation and real-time RT-PCR analysis

Bone marrow cells from ten donor mice were collected and sorted using the marker genes indicated, and VEGF gene ablation was induced as described. Total RNA from different subpopulations expressing the marker genes indicated were isolated using the STAT 60 method (TEL-TEST 'B'), and purified on RNeasy Quick spin columns (Qiagen). Probe primer sequences for murine VEGF, VEGFR-1 and VEGFR-2, real-time RT-PCR conditions and data analysis were as described previously²⁸. A total of 50 ng of total RNA was used per RT-PCR reaction. Genomic DNA from bone marrow was isolated and analysed by Taqman (real time) PCR or by Southern blotting as described previously⁸.

Received 13 February; accepted 11 April 2002; doi:10.1038/nature00821.

1. Ferrara, N. *et al.* Heterozygous embryonic lethality induced by targeted inactivation of the VEGF gene. *Nature* **380**, 439–442 (1996).
2. Carmeliet, P. *et al.* Abnormal blood vessel development and lethality in embryos lacking a single VEGF allele. *Nature* **380**, 435–439 (1996).
3. Janowska-Wieczorek, A., Majka, M., Ratajczak, J. & Ratajczak, M. Z. Autocrine/paracrine mechanisms in human hematopoiesis. *Stem Cells* **19**, 99–107 (2001).
4. Bautz, F., Rafii, S., Kanz, L. & Mohle, R. Expression and secretion of vascular endothelial growth factor-A by cytokine-stimulated hematopoietic progenitor cells. Possible role in the hematopoietic

microenvironment. *Exp. Hematol.* **28**, 700–706 (2000).

5. Shalaby, F. *et al.* Failure of blood-island formation and vasculogenesis in Flk-1-deficient mice. *Nature* **376**, 62–66 (1995).
6. Kabrun, N. *et al.* Flk-1 expression defines a population of early embryonic hematopoietic precursors. *Development* **124**, 2039–2048 (1997).
7. Radtke, F. *et al.* Deficient T cell fate specification in mice with an induced inactivation of Notch1. *Immunity* **10**, 547–558 (1999).
8. Gerber, H. P. *et al.* VEGF is required for growth and survival in neonatal mice. *Development* **126**, 1149–1159 (1999).
9. Solar, G. P. *et al.* Role of c-mpl in early hematopoiesis. *Blood* **92**, 4–10 (1998).
10. Austin, T. W., Solar, G. P., Ziegler, F. C., Liem, L. & Matthews, W. A role for the Wnt gene family in hematopoiesis: expansion of multilineage progenitor cells. *Blood* **89**, 3624–3635 (1997).
11. Browder, T. M., Abrams, J. S., Wong, P. M. & Nienhuis, A. W. Mechanism of autocrine stimulation in hematopoietic cells producing interleukin-3 after retrovirus-mediated gene transfer. *Mol. Cell. Biol.* **9**, 204–213 (1989).
12. Lang, R. A., Metcalf, D., Gough, N. M., Dunn, A. R. & Gonda, T. J. Expression of a hemopoietic growth factor cDNA in a factor-dependent cell line results in autonomous growth and tumorigenicity. *Cell* **43**, 531–542 (1985).
13. Wedge, S. R. *et al.* ZD4190: an orally active inhibitor of vascular endothelial growth factor signalling with broad-spectrum antitumor efficacy. *Cancer Res.* **60**, 970–975 (2000).
14. Sun, L. *et al.* Design, synthesis, and evaluations of substituted 3-[(3- or 4-carboxyethyl)pyrrol-2-yl)methylidene]indolin-2-ones as inhibitors of VEGF, FGF, and PDGF receptor tyrosine kinases. *J. Med. Chem.* **42**, 5120–5130 (1999).
15. Gille, H. *et al.* Analysis of biological effects and signalling properties of Flt-1 and KDR: A reassessment using novel highly receptor-specific VEGF mutants. *J. Biol. Chem.* **275**, 3222–3230 (2000).
16. Park, J. E., Chen, H. H., Winer, J., Houck, K. A. & Ferrara, N. Placenta growth factor. Potentiation of vascular endothelial growth factor bioactivity, in vitro and in vivo, and high affinity binding to Flt-1 but not to Flk-1/KDR. *J. Biol. Chem.* **269**, 25646–25654 (1994).
17. Zeng, H., Dvorak, H. F. & Mukhopadhyay, D. Vascular permeability factor (VPF)/vascular endothelial growth factor (VEGF) receptor-1 down-modulates VPF/VEGF receptor-2-mediated endothelial cell proliferation, but not migration, through phosphatidylinositol 3-kinase-dependent pathways. *J. Biol. Chem.* **276**, 26969–26979 (2001).
18. Barleon, B. *et al.* Migration of human monocytes in response to vascular endothelial growth factor (VEGF) is mediated via the VEGF receptor flt-1. *Blood* **87**, 3336–3343 (1996).
19. Lyden, D. *et al.* Impaired recruitment of bone-marrow-derived endothelial and hematopoietic precursor cells blocks tumour angiogenesis and growth. *Nature Med.* **7**, 1194–1201 (2001).
20. Gerber, H. P. *et al.* Vascular endothelial growth factor regulates endothelial cell survival through the phosphatidylinositol 3'-kinase/Akt signal transduction pathway. Requirement for Flk-1/KDR activation. *J. Biol. Chem.* **273**, 30336–30343 (1998).
21. Behringer, D., Kresin, V., Henschler, R., Mertelsmann, R. & Lindemann, A. Cytokine and chemokine production by CD34⁺ haemopoietic progenitor cells: detection in single cells. *Br. J. Haematol.* **97**, 9–14 (1997).
22. Bellamy, W. T. *et al.* Vascular endothelial cell growth factor is an autocrine promoter of abnormal localized immature myeloid precursors and leukemia progenitor formation in myelodysplastic syndromes. *Blood* **97**, 1427–1434 (2001).
23. Fiedler, W. *et al.* Vascular endothelial growth factor, a possible paracrine growth factor in human acute myeloid leukemia. *Blood* **89**, 1870–1875 (1997).
24. Dias, S. *et al.* Inhibition of both paracrine and autocrine VEGF/VEGFR-2 signalling pathways is essential to induce long-term remission of xenotransplanted human leukemias. *Proc. Natl. Acad. Sci. USA* **98**, 10857–10862 (2001).
25. Taraseviciene-Stewart, L. *et al.* Inhibition of the VEGF receptor 2 combined with chronic hypoxia causes cell death-dependent pulmonary endothelial cell proliferation and severe pulmonary hypertension. *FASEB J.* **15**, 427–438 (2001).
26. Kasahara, Y. *et al.* Inhibition of VEGF receptors causes lung cell apoptosis and emphysema. *J. Clin. Invest.* **106**, 1311–1319 (2000).
27. Ryan, A. M. *et al.* Preclinical safety evaluation of rhuMabVEGF, an antiangiogenic humanized monoclonal antibody. *Toxicol. Pathol.* **27**, 78–86 (1999).
28. Gerber, H. P., Kowalski, J., Sherman, D., Eberhard, D. A. & Ferrara, N. Complete inhibition of rhabdomyosarcoma xenograft growth and neovascularization requires blockade of both tumour and host vascular endothelial growth factor. *Cancer Res.* **60**, 6253–6258 (2000).
29. Hennequin, L. E. *et al.* Design and structure-activity relationship of a new class of potent VEGF receptor tyrosine kinase inhibitors. *J. Med. Chem.* **42**, 5369–5389 (1999).
30. Smolich, B. D. *et al.* The antiangiogenic protein kinase inhibitors SU5416 and SU6668 inhibit the SCF receptor (c-kit) in a human myeloid leukemia cell line and in acute myeloid leukemia blasts. *Blood* **97**, 1413–1421 (2001).

Acknowledgements

We thank E. Humke and F. Radtke for discussions and advice; and L. Lasky, F. de Sauvage and D. Eaton for critical review of the manuscript and comments. We also thank M. Kelly, K. Broderick and the Genentech animal facility for animal care and breeding; L. Thom for tail DNA isolation and genotype analysis; and B. Li and B. Moffat for VEGF mutants. We also thank P. Schow and C. Grimmer for FACS analysis; G. Hatami, L. Leong and L. Cabote for blood cell analysis; and M. Van Hoy, E. Pegg, D. Eberhard and K. Hillan for pathology support.

Competing interests statement

The authors declare competing financial interests: details accompany the paper on *Nature's* website (<http://www.nature.com/nature>).

Correspondence and requests for materials should be addressed to H.-P.G. (e-mail: gerberhp@gene.com).

A plant receptor-like kinase required for both bacterial and fungal symbiosis

Silke Stracke*, Catherine Kistner*, Satoko Yoshida*, Lonneke Mulder*, Shusei Sato†, Takakazu Kaneko†, Satoshi Tabata†, Niels Sandal‡, Jens Stougaard‡, Krzysztof Szczypkowski§ & Martin Parniske*

* The Sainsbury Laboratory, John Innes Centre, Colney Lane, Norwich NR4 7UH, UK

† Kazusa DNA Research Institute, 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
‡ Laboratory of Gene Expression, IMSB, University of Aarhus, Gustav Wieds Vej 10, DK-8000 Aarhus C, Denmark

§ Agriculture and Agri-Food Canada, Southern Crop Protection and Food Research Centre, 1391 Sandford Street, London, Ontario N5V 4T3, Canada

Most higher plant species can enter a root symbiosis with arbuscular mycorrhizal fungi, in which plant carbon is traded for fungal phosphate^{1,2}. This is an ancient symbiosis, which has been detected in fossils of early land plants³. In contrast, the nitrogen-fixing root nodule symbioses of plants with bacteria evolved more recently, and are phylogenetically restricted to the rosoid I clade of plants⁴. Both symbioses rely on partially overlapping genetic programmes^{5,6}. We have identified the molecular basis for this convergence by cloning orthologous *SYMRK* ('symbiosis receptor-like kinase') genes from *Lotus* and pea, which are required for both fungal and bacterial recognition. *SYMRK* is predicted to have a signal peptide, an extracellular domain comprising leucine-rich repeats, a transmembrane and an intracellular protein kinase domain. *Lotus SYMRK* is required for a symbiotic signal transduction pathway leading from the perception of microbial signal molecules to rapid symbiosis-related gene activation. The perception of symbiotic fungi and bacteria is mediated by at least one common signalling component, which could have been recruited during the evolution of root nodule symbioses from the already existing arbuscular mycorrhiza symbiosis.

In root nodule symbioses, bacteria of the genera *Frankia* and *Rhizobium* are hosted within a newly formed plant organ, the root nodule, and fix atmospheric nitrogen, enabling plants to grow without added nitrate or ammonia. The majority of higher plant species, including major crops such as wheat and rice, can only engage in arbuscular mycorrhiza but not in root nodule symbioses. Legume mutants have been described which are defective in arbuscular mycorrhiza and root nodule symbioses, thus indicating that there are common genetic components to the two symbioses^{1,5}. In order to define this genetic overlap at the molecular level, we isolated symbiotic mutants from the model legume *Lotus japonicus*⁶⁻⁹.

Lotus SYMRK mutants were unable to form root nodules and arbuscular mycorrhiza. Their phenotype implicated the affected gene in early stages of symbiotic infection. On wild-type *L. japonicus* roots, intracellular infection by fungal hyphae progresses from root epidermal cells to cortical cells in which arbuscules develop⁹. The infection by arbuscular mycorrhiza fungi is accompanied by the formation of an intracellular accommodation structure by the plant host cells⁹. On roots of *Lotus SYMRK* mutants, most fungal infection attempts are arrested in epidermal cells within which hyphae frequently exhibit aberrant swellings and deformations⁶, indicating that the mutant epidermal cells are unable to support fungal infection.

Bacterial infection of *L. japonicus* roots by *Mesorhizobium loti* is initiated in root hairs; wild-type root hairs respond with branching and curling, which results in entrapment of a bacterial micro-colony

in an infection pocket, where the infection thread, an intracellular infection structure, is initiated (Fig. 1a). In all tested alleles of *Lotus SYMRK*, root hairs responded with exaggerated swelling and branching when inoculated with *M. loti*, but not with curling. Neither entrapment of bacteria in infection pockets nor infection threads were observed (Fig. 1b), indicating that *Lotus SYMRK* is required for infection thread initiation. As the development of an intracellular infection structure is impaired in the interaction with both rhizobia and fungi, we conclude that *Lotus SYMRK* is required for a common developmental step leading to the intracellular accommodation of symbiotic microorganisms.

Rhizobia synthesize so called Nod factors (NF), symbiotic signalling compounds identified as lipochitin oligosaccharides, that induce root hair deformation¹⁰. When inoculated with a *M. loti nodC::Tn5* mutant, defective in the production of NFs, root hairs of wild-type and mutant roots did not deform (Fig. 1c, d), indicating that NF is required for this root hair response. As mutant root hairs deform in an NF-dependent fashion, we conclude that NF-dependent signalling leading to root hair deformation is independent of *Lotus SYMRK*.

After inoculation with *M. loti*, only a subset of root hairs within a susceptible region behind the tip of wild-type roots showed morphological responses. In contrast, almost all root hairs of *Lotus SYMRK* mutants deformed (Fig. 1b). Such an increase in the number of responding root hairs has been observed in other infection-defective legume mutants, and is interpreted as evidence for a mechanism which negatively regulates root cell competence in wild-type plants¹¹.

We studied the expression of a symbiosis-activated gene to establish the position of *SYMRK* in symbiotic signalling. Leghaemoglobin (LB), although classically ascribed the role of an oxygen buffer in the mature nodule, might have additional roles in the early infection process as *LB* gene activation has been detected within hours after NF treatment¹². In *L. japonicus* roots, *LB* is activated 24 and 48 hours after inoculation with *M. loti* wild-type but not after inoculation with a *M. loti nodC::Tn5* mutant unable to produce NF

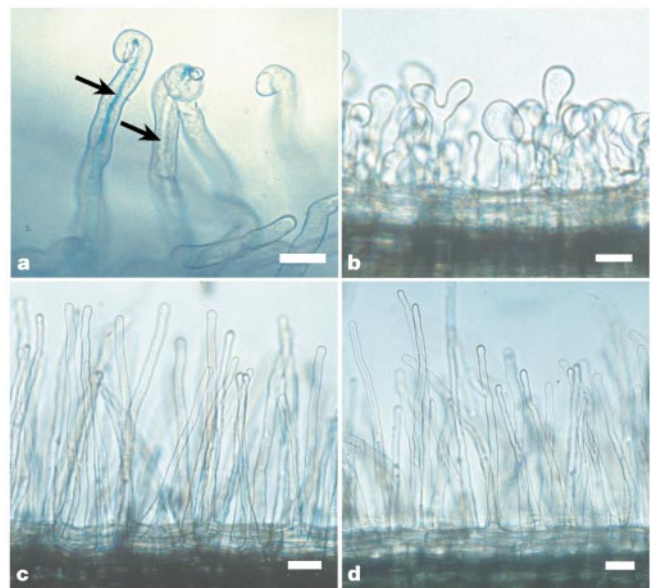


Figure 1 Root hair responses of wild-type and a *Lotus SYMRK* mutant to bacterial inoculation. Light micrographs of **a**, infection threads (arrows) in root hairs of wild-type plants and **b**, root hair deformation, but no infection thread formation in *Lotus SYMRK* mutant *cac41.5*, both inoculated with *M. loti* R7A. **c**, Straight root hairs in wild-type plants and **d**, *Lotus SYMRK* mutant *cac41.5*, both inoculated with *M. loti* R7AC2, a *nodC::Tn5* mutant. Scale bars: **a**, **b**, 50 μm; **c**, **d**, 100 μm.

(Fig. 2). Moreover, *LB* is activated in roots treated with NF but not in water-treated control roots (Fig. 2). These results demonstrate that NF is necessary and sufficient for early *LB* gene activation. Activation of the *LB* gene is absent in *Lotus SYMRK* mutants *cac41.5* (Fig. 2) and EMS61 (data not shown), indicating that SYMRK is required for NF-mediated induction of *LB* gene expression. These results demonstrate that *Lotus SYMRK* has a role in early signal transduction between NF perception and activation of the *LB* gene.

Lotus SYMRK was isolated by map-based cloning and could be positioned on the bottom arm of chromosome 2 between the markers *AUX* and *SHMT* (Fig. 3a). A contig of transformation-competent artificial chromosome (TAC) clones was assembled that comprised the *Lotus SYMRK* region, including markers flanking the gene on either side (Fig. 3b). The TAC contig was entirely sequenced in the context of the *L. japonicus* genome sequencing project¹³, and the *Lotus SYMRK* gene could be delimited to a region of less than 120 kilobases (kb) using recombination events (Fig. 3b). Eighteen putative genes were predicted within this target interval. Upon BLAST searches of the predicted coding sequences, ten resulted in either weak homologies or in homologies to unknown or putative proteins. The remaining eight genes had homology to an oligopeptidase, a receptor-like kinase (RLK), an ABC transporter, an aminoacyl-tRNA ligase, a β -xylosidase, two RNA polymerase subunits and one retroelement. The RLK appeared to be an obvious candidate for a signal transduction component, a function compatible with the *Lotus SYMRK* mutant phenotype. This also fitted the suggestion that RLKs may be involved in early symbiotic signal transduction processes¹⁴. Therefore, a complementary DNA of the RLK gene was amplified by polymerase chain reaction (PCR), sequenced and 15 exons were detected upon alignment of the cDNA with the genomic sequence (Fig. 3c).

The conceptual open reading frame starts with two consecutive AUGs. The first potential initiator codon is preceded by an in-frame stop codon located 18 base pairs (bp) upstream. The determined cDNA sequence carries a full-length open reading frame of 2,769 nucleotides coding for a protein of 923 amino acids with a predicted relative molecular mass of 103,000 (M_r 103K). The extent of the 3' untranslated region and a position of the polyA tail was determined from a matching expressed sequence tag (EST) cDNA clone (MWL068h06). Three independent mutations were found within the RLK gene upon analysis of three mutant alleles (Fig. 3c). Mutant EMS61 was generated by EMS mutagenesis⁷ and suffered from a nonsense mutation in the kinase domain leading to a predicted truncated protein of 805 amino acids (Fig. 4). Mutants 282-287 and *cac41.5* resulted from an attempt to generate T-DNA or transposon-tagged mutants of *L. japonicus*⁸, but none of the alleles were found to

be tagged. Mutants 282-287 and *cac41.5* both carry large insertions of approximately 4.9 kb in exon 3, and 5.8 kb in exon 4, respectively (Fig. 3c). These insertions possibly originated from endogenous transposons. More than one mis-spliced messenger RNA product

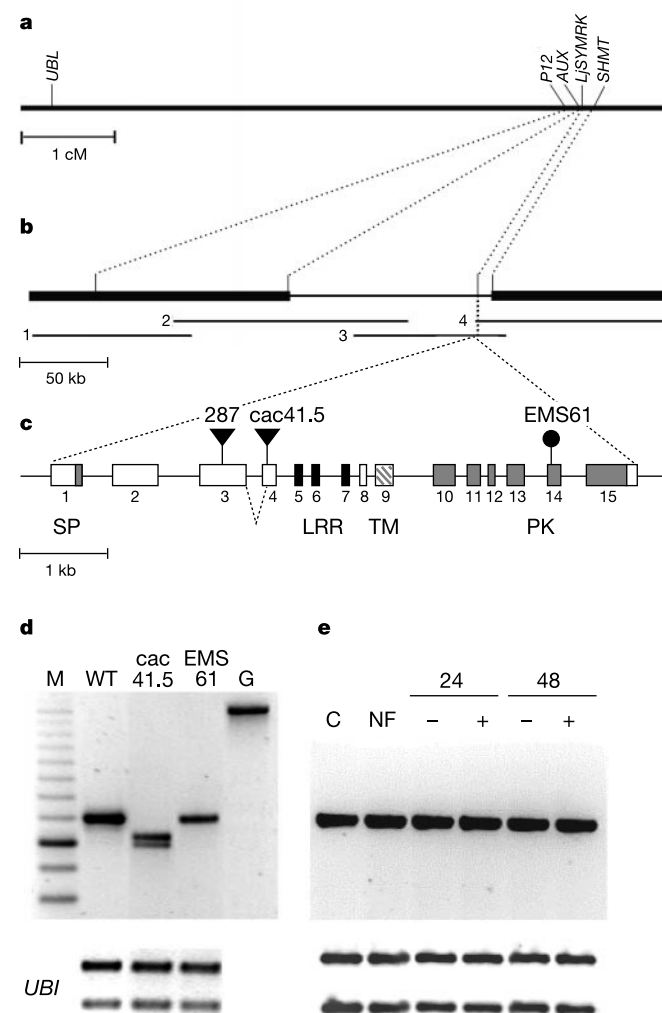


Figure 3 Positional cloning of *Lotus SYMRK*. **a**, Genetic map of the *Lotus SYMRK* region on the long arm of *L. japonicus* chromosome 2. *UBL*, *AUX* and *SHMT* refer to an ubiquitin-like gene, an auxin-inducible gene and a serine-hydroxymethyltransferase gene, respectively. *P12* is an AFLP marker. **b**, Physical map of a TAC contig comprising *Lotus SYMRK*. The *Lotus SYMRK* target region could be delimited by recombination events between *Lotus SYMRK* and the flanking markers *AUX* and *SHMT*. The resulting target interval is indicated by a thin line flanked by bold lines. Extensions of TAC sequences are indicated below the map. Numbers 1, 2, 3 and 4 refer to TACs *Lotus* T10N22, *Lotus* T01N11, *Lotus* T11E23 and *Lotus* T17H01 respectively. **c**, Intron-exon structure of the *Lotus SYMRK* gene and positions of the coded protein domains. SP, signal peptide; LRR, leucine-rich repeats; TM, trans-membrane domain; PK, protein kinase. The positions of insertions in mutants 282-287 (287) and *cac41.5*, and the position of a nonsense mutation in mutant EMS61 are indicated. The dotted lines leading from the end of exon 3 into exon 4 indicate the position of two mis-splicing events identified in mutant *cac41.5*. **d**, *Lotus SYMRK* cDNA analysis by RT-PCR in roots of wild-type ('Gifu') (WT) and mutants *cac41.5* and EMS61, using a forward primer in exon 2 and a reverse primer in exon 5. Two alternative splice products could be detected in roots of mutant *cac41.5*, which were 76 and 105 bases shorter than the wild-type cDNA of 902 bp, whereas a wild-type sized PCR product was amplified from mutant EMS61. All cDNA samples were harvested 48 h after inoculation with *M. loti* R7A. Wild-type genomic DNA (1,625 bp, G) was amplified for comparison. M, 100-bp ladder. **e**, *Lotus SYMRK* is expressed constitutively in roots. Expression of the *Lotus SYMRK* gene in wild-type roots treated as indicated in legend of Fig. 2. **d**, **e**, Polyubiquitin (*UBI*) cDNA fragments were amplified as a control for equal amounts of cDNA template.

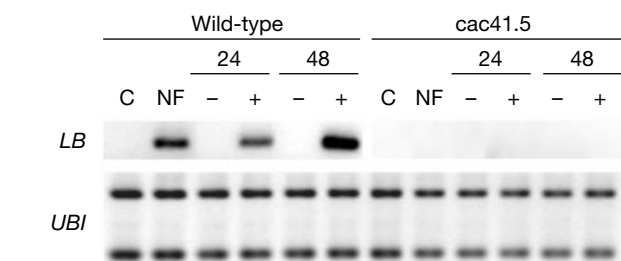


Figure 2 NF-induced gene activation is *Lotus SYMRK*-dependent. Gene expression in *L. japonicus* Gifu (wild-type) and *Lotus SYMRK* mutant (*cac41.5*) roots analysed by RT-PCR. Roots were treated with water for 48 h for control (C) or with Nod factor for 18 h (NF), or inoculated with *M. loti* strain R7A (+) or strain R7AC2 carrying a *nodC::Tn5* mutation (-) and harvested 24 or 48 h post inoculation. Expression of a leghaemoglobin gene (*LB*) in roots treated as indicated. Polyubiquitin cDNA fragments (*UBI*) were amplified as a control for equal amounts of template in each sample.

could be detected in mutant cac41.5 (Fig. 3c, d), probably a consequence of the insertion. Based on the physical and genetical map location and identification of mutations in three independent alleles, we concluded that the *Lotus SYMRK* gene is identical with the predicted *RLK* gene. As SYMRK is a potential signalling component, we investigated whether its mRNA levels would change upon signal perception. Using PCR after reverse transcription of RNA (RT-PCR), *Lotus SYMRK* was found to be constitutively expressed in roots and the amount of message did not change upon treatment with NF or 24 h and 48 h after inoculation with *M. loti* (Fig. 3e).

The predicted protein sequence coded for by the *Lotus SYMRK* gene comprises the classical hallmarks of an RLK; a signal peptide, an extracellular domain, a transmembrane domain and an intracellular protein kinase domain (Figs 3c and 4). The protein kinase domain is highly homologous to that of other plant RLKs. Irrespective of whether the kinase domain or the extracellular domain of *Lotus SYMRK* were used for BLAST database searches, the closest homologues in *Arabidopsis* were the two RLKs Atlg67720 and At2g37050, which together with its similar overall domain and intron-exon structure, places SYMRK in the LRR class of RLKs¹⁵.

At the amino terminus, a hydrophobic segment of 29 amino acids presumably acts as a signal peptide (Fig. 4). Three leucine-rich repeats (LRRs) of the extracellular type are found in the predicted extracellular domain of SYMRK. LRRs have been shown to mediate specific protein-protein interactions¹⁶. The region coding for the LRRs has a modular intron-exon structure (Figs 3c and 4). The three LRRs of 24, 23 and 24 amino acids in length are encoded by 3 exons of 72, 69 and 72 nucleotides respectively, with each exon encoding one LRR unit. Introns interrupt the LRR coding sequences at conserved positions, after the second nucleotide of the codon for the first leucine (L) of the predicted β -strand- β -turn region xxLxLxx of the LRR (Fig. 4). This intron position is conserved in a number of other plant and animal genes coding for extracellular LRRs suggesting a common single LRR-exon ancestor¹⁶. It is also characteristic for members of the LRR1 and LRR2 classes of plant RLKs¹⁵. A fourth exon of 72 nucleotides possibly also evolved by duplication of an ancestral LRR exon but lacks structurally important leucines (Fig. 4).

The position of *Lotus SYMRK* in an early symbiotic signalling pathway is reminiscent of pea and *Medicago truncatula* genes required for both root symbioses^{17–19}. *Lotus SYMRK* and the pea *SYM19* mutants exhibited a similar phenotype. Moreover, the marker SHMT which is less than 8 kb away from *Lotus SYMRK* (Fig. 3a, b) is also closely linked to the *SYM19* gene in pea^{20,21}. We therefore investigated whether the genes are orthologous. EST database searches using *Lotus SYMRK* cDNA sequence revealed identical ESTs from *L. japonicus* and highly homologous ESTs from *M. truncatula* and soybean, which are likely to represent orthologues of *Lotus SYMRK*. A pea homologue to *Lotus SYMRK* was amplified using stretches of sequence conservation between the available legume sequences for primer design. A cDNA sequence was obtained, and the coding region was respectively 85.7% and 82.8% identical at the nucleotide and peptide level to *Lotus SYMRK*. cDNA fragments were amplified from wild-type cultivar 'Frisson' and two symbiotic mutants of *SYM19*. In mutant P4⁵, a point mutation in subdomain I was detected in the consensus motif of the glycine-rich loop GXGXXGXV, an ATP anchor present in nearly all protein kinases. The codon GGA which codes for the second glycine (G), was mutated to GAA, which changed the predicted protein sequence GEGGFGSV into GEEGFGSV. In mutant P55⁵, the codon GAT coding for the aspartate residue (D) in the conserved DFG motif of subdomain VII was mutated to AAT coding for an asparagine. Both the mutated glycine and aspartate have been implicated in ATP-binding and these mutations are therefore likely to affect catalytic activity. Based on similar genomic environment, similar mutant phenotype, strong sequence homology and sequence analysis of two mutant alleles, we concluded that pea *SYM19* and *Lotus SYMRK* are orthologous genes.

Here we describe a component of a plant signalling pathway leading to intracellular accommodation of both bacterial and fungal endosymbionts. When applied to legume root hairs, NFs induce within minutes rhythmic fluctuations in intracellular calcium also known as calcium-spiking²². The pea gene *SYM19* is required for the induction of calcium-spiking¹⁸, which places *SYMRK* upstream of one of the earliest detectable root hair responses. SYMRK is a receptor-like kinase and acts between the recognition of microbial signalling molecules and the activation of calcium-spiking and symbiosis-related (*LB*) gene expression. The RLK structure of SYMRK favours a model in which the extracellular domain is involved in the perception of an extracellular ligand, and the signal is transduced through the intracellular kinase domain. Whether the LRRs of SYMRK are required for the direct interaction with a signalling ligand, or whether they have a role in the assembly of a receptor complex, has yet to be determined. It is possible that the bacterial and fungal signal molecules are sufficiently similar to be recognized by the same receptor, or alternatively, additional specific components are involved which perceive either the rhizobial NF or an as yet hypothetical signal emanating from the arbuscular mycorrhiza fungus. □

Methods

Plant material

The pea mutants P4 and P55 of cultivar Frisson have been described previously²³, and belong to the *SYM19* complementation group²⁰. All *L. japonicus* mutants used in this study have been generated in ecotype 'Gifu'. Allelism tests revealed that the mutants EMS61 generated by EMS-mutagenesis⁷, and the mutants 282-287 (ref. 8) and cac41.5 (J.S., unpublished results) generated through a transposon/T-DNA-tagging attempt, belong to the previously defined *Lotus SYM2* complementation group⁸. To reflect the predicted protein function we have renamed *Lotus SYM2* and refer to it as *Lotus SYMRK* ('symbiosis receptor-like kinase').

L. japonicus seedlings were grown for two weeks in the presence of 0.1 mg l⁻¹ 1- α -(2-aminoethoxyvinyl)-glycine²⁴ and inoculated with *M. loti* R7A or strain R7AC2 (*nodC::Tn5* mutant; J. T. Sullivan and C. W. Ronson, personal communication) (200 μ l per root, $A_{600} = 0.1$), with NF (200 μ l per root, 10⁻⁷ M) prepared from strain R7A by reverse-phase liquid chromatography²⁵, or treated with sterile water as a control. The NF preparation contains as the predominant component a N-acetylglucosamine pentasaccharide of which the non-reducing residue is N-methylated and N-acetylated with *cis*-vaccenic acid (C18:1)

MMELPATRILSQAVTCFLCLYIFIGSASA SP
TEG+FESIAACADLNYTDPLTLLNYTDDYTFWSDKRSCKRIPETELNRNSNENVR
FDIDEGKRCYNLPTIKNGVYLIRGTFFPDSLNSSFNASIGVTQLGAVRSSRLQDLE
IEGVFRATKDYIDFCLLKEGVYFPISQLELRPSPEEYLQDFPTSVLKLISRNLLGD
TKDDIR+FPVDQSDRIWKASSISSAVPLSSNVSNVDLNAVTPPLTVLQATLDP
ERLEFIHTDLETDYGYRVFLYFLELDRLTLAGQGVFDIYVNSEIKKESFDVLAGE
SNRYVDLDSASGSLNVTLVKASKSEFGPLLNAYEILQVRWIEETNQTD+VGVI
QKMRRELLLNQSGNRALESWSGDPICLLPWKGIACDGSNGSS 404
[LXXLXLXNLXGXIF] cons
VITKL+DLSSNKLGLIPSSIAEMT LRR1
NLETL+NISHNSFDGSVP SFPLSS LRR2
LLISV+DLSYNDLMGKLPEISVKLP LRR3
HLKSL+YFGCNEHMSPEDPANMNSS
LINTD+YGRCKGKESRFGQ 517
VIVIGAITCGSLITLAFGVLFV TM
CRYRQKLIPEWEGFAGKKYPMET+NIIFSLPSKDDFFIKSVSIQAFTELEYIEVAT 593
ERYKTLIGEGGFGSVYRGTLDNGQEVAVKVSATSTQGTREFDNE+LNLSSAIQHE PK
NLVPLLGYCNESDQILVYFPFMSNGSLQDRLY+GEPAKRKILDWPTRLSIALGAAR
+GLAYLHTFPGRSVIHRDIKSSNILLDHSMCAKVAQDFGFSKYAPQEGDSYVSLEVR
GTAGYLDPE+YYKTQQLSEKSDVFSFGVLLIIVSGREPLNIKRPTE[MSLVEW+A
TPYIRGSKVDEIVDPGKIGGYHAEAMWRVVEVALQCLEPFSTYRPSMVAIVRELD
ALTIENNASEYMKSIDSLGGSNRYSVIEIKRVLPTSTTAESTITTTQSLSHPPQR* 923

Figure 4 Features of the *Lotus SYMRK* amino-acid sequence. SP, signal peptide; cons, consensus motif for extracellular leucine-rich repeats¹⁶; LRR 1–3, leucine-rich repeats 1–3; TM, trans-membrane domain; PK, protein kinase domain; + indicates approximate relative positions of introns. The position of the nonsense mutation in mutant EMS61 is indicated by a box.

and carries a carbamoyl group. The reducing terminal residue is substituted with 4-O-acetylucose. This is consistent with published *M. loti* NF structure for R7A²⁴. A derivative in which the terminal residue is substituted with fucose was identified as a minor constituent. At set times after inoculation, root tips (~3 mm) were discarded and susceptible zones (3–4 cm) of 14 roots each were collected in liquid nitrogen and used for RNA preparation. For root hair observations, seedlings were inoculated with *M. loti* four days after germination. To visualize infection threads, roots were inoculated with a *M. loti* R7A derivative carrying pXLGD4 constitutively expressing a *lacZ* reporter gene¹¹. Entire root systems of at least 10 seedlings were inspected per allele, 14 and 21 d post inoculation, after staining for β -galactosidase activity.

RT-PCR

To compare cDNA concentrations in different samples, fragments of a *L. japonicus* polyubiquitin cDNA (GenBank accession number AW719307) were PCR-amplified (small fragment 227 bp) with primers fwd P40079 CTACAACTTCAGAGGAGTCCA and rev P40080 CACAGGCCAGAGAGGCCACAACA (22 cycles: 94 °C for 15 s, 58 °C for 30 s, 72 °C for 50 s). Amplification was exponential up to 25 cycles. A 257-bp fragment of a leghaemoglobin cDNA (AB042718), was PCR-amplified using primers fwd P40183 GTTCTACACCGTTATATTGGAGATAG and rev P41008 GCAGGCTTCTTTAACCACCACG (both spanning exon/intron borders) (40 cycles: 94 °C for 15 s, 59 °C for 30 s, 72 °C for 40 s). For (RT)-PCR of *Lotus SYMRK* (Fig. 3e) primers fwd P60243 TGCAGTGAGATCATCCAGGCTC and rev P60244 TCTAAGTTGGTCATCTCAGCAATGCTGG were used. (35 cycles: 94 °C for 15 s, 65 °C for 30 s, 72 °C for 60 s). PCR fragments were analysed by agarose gel electrophoresis and sequencing.

Received 23 January; accepted 23 April 2002; doi:10.1038/nature00841.

- Harrison, M. J. Molecular and cellular aspects of the arbuscular mycorrhizal symbiosis. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **50**, 361–389 (1999).
- Rausch, C. et al. A phosphate transporter expressed in arbuscule-containing cells in potato. *Nature* **414**, 462–470 (2001).
- Remy, W., Taylor, T. N., Hass, H. & Kerp, H. Four hundred-million-year-old vesicular arbuscular mycorrhizae. *Proc. Natl Acad. Sci. USA* **91**, 11841–11843 (1994).
- Soltis, D. E. et al. Chloroplast gene sequence data suggest a single origin of the predisposition for symbiotic nitrogen-fixation in angiosperms. *Proc. Natl Acad. Sci. USA* **92**, 2647–2651 (1995).
- Duc, G., Trouvelot, A., Gianinazzi-Pearson, V. & Gianinazzi, S. First report of non-mycorrhizal plant mutants (myc⁻) obtained in pea (*Pisum sativum* L.) and fababean (*Vicia faba* L.). *Plant Sci.* **60**, 215–222 (1989).
- Wegel, E., Schausser, L., Sandal, N., Stougaard, J. & Parniske, M. Mycorrhiza mutants of *Lotus japonicus* define genetically independent steps during symbiotic infection. *Mol. Plant-Microbe Interact.* **11**, 933–936 (1998).
- Szczygłowski, K. et al. Nodule organogenesis and symbiotic mutants of the model legume *Lotus japonicus*. *Mol. Plant-Microbe Interact.* **11**, 684–697 (1998).
- Schauser, L. et al. Symbiotic mutants deficient in nodule establishment identified after T-DNA transformation of *Lotus japonicus*. *Mol. Gen. Genet.* **259**, 414–423 (1998).
- Bonfante, P. et al. The *Lotus japonicus* *LjSym4* gene is required for the successful symbiotic infection of root epidermal cells. *Mol. Plant-Microbe Interact.* **13**, 1109–1120 (2000).
- Downie, J. A. & Walker, S. A. Plant responses to nodulation factors. *Curr. Opin. Plant Biol.* **2**, 483–489 (1999).
- Schauser, L., Roussis, A., Stiller, J. & Stougaard, J. A plant regulator controlling development of symbiotic root nodules. *Nature* **402**, 191–195 (1999).
- Heidstra, R., Nilsen, G., Martinez-Abarca, F., van Kammen, A. & Bisseling, T. Nod factor-induced expression of leghemoglobin to study the mechanism of NH₄NO₃ inhibition on root hair deformation. *Mol. Plant-Microbe Interact.* **10**, 215–220 (1997).
- Sato, S. et al. Structural analysis of a *Lotus japonicus* genome. I. Sequence features and mapping of fifty-six TAC clones which cover 5.4 Mb of the genome. *DNA Res.* **8**, 311–318 (2001).
- Cullimore, J. V., Ranjeva, R. & Bono, J. J. Perception of lipo-chitoooligosaccharidic Nod factors in legumes. *Trends Plant Sci.* **6**, 24–30 (2001).
- Shiu, S. H. & Blecker, A. B. Receptor-like kinases from *Arabidopsis* form a monophyletic gene family related to animal receptor kinases. *Proc. Natl Acad. Sci. USA* **98**, 10763–10768 (2001).
- Jones, D. A. & Jones, J. D. G. The role of leucine-rich repeat proteins in plant defences. *Adv. Bot. Res.* **24**, 89–167 (1997).
- Catoira, R. et al. Four genes of *Medicago truncatula* controlling components of a nod factor transduction pathway. *Plant Cell* **12**, 1647–1665 (2000).
- Walker, S. A., Viprey, V. & Downie, J. A. Dissection of nodulation signalling using pea mutants defective for calcium-spiking induced by Nod-factors and chitin oligomers. *Proc. Natl Acad. Sci. USA* **97**, 13413–13418 (2000).
- Wais, R. J. et al. Genetic analysis of calcium spiking responses in nodulation mutants of *Medicago truncatula*. *Proc. Natl Acad. Sci. USA* **97**, 13407–13412 (2000).
- Schneider, A. et al. Genetic mapping and functional analysis of a nodulation-defective mutant (*sym19*) of pea (*Pisum sativum* L.). *Mol. Gen. Genet.* **262**, 1–11 (1999).
- Turner, S. R., Hellens, R., Ireland, R., Ellis, N. & Rawsthorne, S. The organization and expression of the genes encoding the mitochondrial glycine decarboxylase complex and serine hydroxymethyltransferase in pea (*Pisum sativum*). *Mol. Gen. Genet.* **236**, 402–408 (1993).
- Ehrhardt, D. W., Wais, R. & Long, S. R. Calcium spiking in plant root hairs responding to *Rhizobium* nodulation signals. *Cell* **85**, 673–681 (1996).
- Duc, G. & Messager, A. Mutagenesis of pea (*Pisum sativum* L.) and the isolation of mutants for nodulation and nitrogen fixation. *Plant Sci.* **60**, 207–213 (1989).
- van Spronsen, P. C., Gronlund, M., Pacios Bras, C., Spaik, H. P. & Kijne, J. W. Cell biological changes of outer cortical root cells in early determinate nodulation. *Mol. Plant-Microbe Interact.* **14**, 839–847 (2001).
- Firmin, J. L., Wilson, K. E., Carlson, R. W., Davies, A. E. & Downie, J. A. Resistance to nodulation of cv. Afghanistan peas is overcome by *nodX*, which mediates an O-acetylation of the *Rhizobium leguminosarum* lipo-oligosaccharide nodulation factor. *Mol. Microbiol.* **10**, 351–360 (1993).

Acknowledgements

We thank A. Downie and D. Bradley for comments on the manuscript, A. Serna-Sanz for supplying nod-factor preparations, C. Pacios Bras and H. Spaik for making available *M. loti* R7A carrying *lacZ*, C. Ronson for providing the *M. loti* NodC mutant, A. Edwards and C. Martin for help with *Pisum* libraries, and V. Viprey for the gift of *Pisum* and *Medicago* cDNA. We thank M.-A. Torres for structure prediction. C.K. was supported by a postdoctoral fellowship from the DFG (Deutsche Forschungsgemeinschaft). Research at the Sainsbury Laboratory is funded by the Gatsby Charitable Foundation.

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to M.P.

(e-mail: martin.parniske@bbsrc.ac.uk). Sequences are deposited at GenBank with the following accession numbers. *Lotus SYMRK*, AF492655; pea *SYM19*, AF491997; *M. truncatula SYMRK*, AF491998. TAC clones: *Lotus* T11E23 (containing the *Lotus SYMRK* gene), AP004579; *Lotus* T01N11, AP004576; *Lotus* T10N22, AP004577; *Lotus* T17H01, AP004578.

A receptor kinase gene regulating symbiotic nodule development

Gabriella Endre, Attila Kereszt, Zoltán Kevei, Sorina Mihacea, Péter Kaló & György B. Kiss

Institute of Genetics, Biological Research Center of the Hungarian Academy of Sciences, H-6701 Szeged, PO Box 521, Hungary

Leguminous plants are able to establish a nitrogen-fixing symbiosis with soil bacteria generally known as rhizobia. Metabolites exuded by the plant root activate the production of a rhizobial signal molecule, the Nod factor, which is essential for symbiotic nodule development^{1,2}. This lipo-chitoooligosaccharide signal is active at femtomolar concentrations, and its structure is correlated with host specificity of symbiosis³, suggesting the involvement of a cognate perception system in the plant host. Here we describe the cloning of a gene from *Medicago sativa* that is essential for Nod-factor perception in alfalfa, and by genetic analogy, in the related legumes *Medicago truncatula* and *Pisum sativum*. The identified 'nodulation receptor kinase', NORK, is predicted to function in the Nod-factor perception/transduction system (the NORK system) that initiates a signal cascade leading to nodulation. The family of 'NORK extracellular-sequence-like' (NSL) genes is broadly distributed in the plant kingdom, although their biological function has not been previously ascribed. We suggest that during the evolution of symbiosis an ancestral NSL system was co-opted for transduction of an external ligand, the rhizobial Nod factor, leading to development of the symbiotic root nodule.

The tetraploid alfalfa non-nodulation mutant MN-1008 (ref. 4) is known to lack all symbiotic responses upon inoculation with compatible *Sinorhizobium meliloti* or treatment with the corresponding cognate Nod factor^{5–7}. But spontaneous nodulation in the absence of *Sinorhizobium*, a Nod-factor-independent phenotype reported so far only on alfalfa⁸, could be detected in the mutant⁹. Consequently, the perception of the extrinsic signal was abolished in MN-1008, but the capacity for nodule initiation and development, as an internal developmental programme of the plant, remained intact. Like several other non-nodulation mutants, MN-1008 is also resistant to colonization by vesicular-arbuscular mycorrhizal fungi (the Myc⁻ phenotype)¹⁰. This genetic intersection between mycorrhizal and rhizobial associations has attracted significant interest because it suggests that symbiotic nitrogen fixation may be derived, in part, from the more ancient and widely distributed mycorrhizal symbiosis¹¹. Thus, determining the molecular basis of mutations

such as in MN-1008 is central to understanding two of the most important symbiotic associations of plants, and their genetic overlap in legumes.

Map-based cloning of the genetic determinant responsible for the non-nodulation trait of *M. sativa* MN-1008 was initiated by mapping the Nod⁻ phenotype in a tetraploid segregating population of alfalfa¹². As this study demonstrated a high level of synteny between the *M. sativa* and *M. truncatula* genomes, a bacterial artificial chromosome (BAC) library of *M. truncatula*¹³ was used to construct a contig spanning the mutation. Two tightly linked restriction-fragment length polymorphism (RFLP) markers¹², SHMT and CAD, were used to isolate primary BAC clones, and a 600-kilobase (600-kb) contig (the 'Nod contig') was assembled based on a combination of chromosome walking from BAC end sequences and restriction endonuclease fingerprinting of the clones. Genetic markers internal to the Nod contig were used to position the Nod⁻ mutation within a fine-structure linkage map (Fig. 1a). Genotypes were scored in a segregating tetraploid *M. sativa* F₂ population (NAB population¹²) with the help of recombinant plants. This analysis revealed correspondence between the *M. truncatula* physical map and the genetic map positions in *M. sativa*. Moreover, two recombinant plants, NAB 4156 and 4443, served to delimit the MN-1008 non-nodulation mutation (*nn1*) to a ~160-kb interval (Fig. 1a, b). BAC clones containing the minimum tiling path of the Nod region (part of BAC 67A11, and the entirety of BACs 2D11 and 28I12; Fig. 1b) were subcloned and sequenced (data available at <http://www.szbk.u-szeged.hu/~alfi/>). A total of 15 open reading frames were identified and putative functions were assigned based on sequence homologies (see Fig. 1b and Supplementary Information). The predicted gene content of the Nod contig shared

limited microsynteny with regions of the *Arabidopsis thaliana* genome (Fig. 1c), but functional information for the syntenic counterparts in *Arabidopsis* did not enhance our ability to identify candidate genes.

The non-nodulation phenotype of MN-1008 could be explained by an alteration in several candidate genes in this region, including genes predicted to code for an ATP-binding cassette (ABC) transporter, a receptor kinase, a MADS-box protein and lectin proteins. We first chose the ABC transporter and receptor kinase, membrane-spanning proteins to analyse for genetic alterations by comparing the sequence of alleles from wild-type nodulating alfalfa and the mutant MN-1008 plant material. To reduce confusion from different F₂ alleles, F₃ individuals that were homozygous throughout the Nod contig region were identified and used for sequencing. Based on the *M. truncatula* BAC sequences, exon-specific primers were designed to amplify both genomic and complementary DNA templates. At least three clones from independent amplifications of overlapping portions of both genes were analysed from the wild-type and mutant backgrounds of *M. sativa*. Based on this analysis, only synonymous changes were detected in the ABC transporter coding genes, whereas an unambiguous mutation, an 'in frame' stop codon, was identified in the receptor kinase gene from the mutant. This mutation was found consistently in independent amplification products from both genomic and cDNA templates of MN-1008 mutant. We have named this receptor kinase gene 'nodulation receptor kinase', *NORK*.

In addition to the analyses of *Medicago* species, we sequenced the candidate *NORK* gene from related legume species *Melilotus alba*, *P. sativum*, *Vicia hirsuta* and *Lotus japonicus*, and in each case the deduced proteins exhibited high overall homology. As shown in Fig.

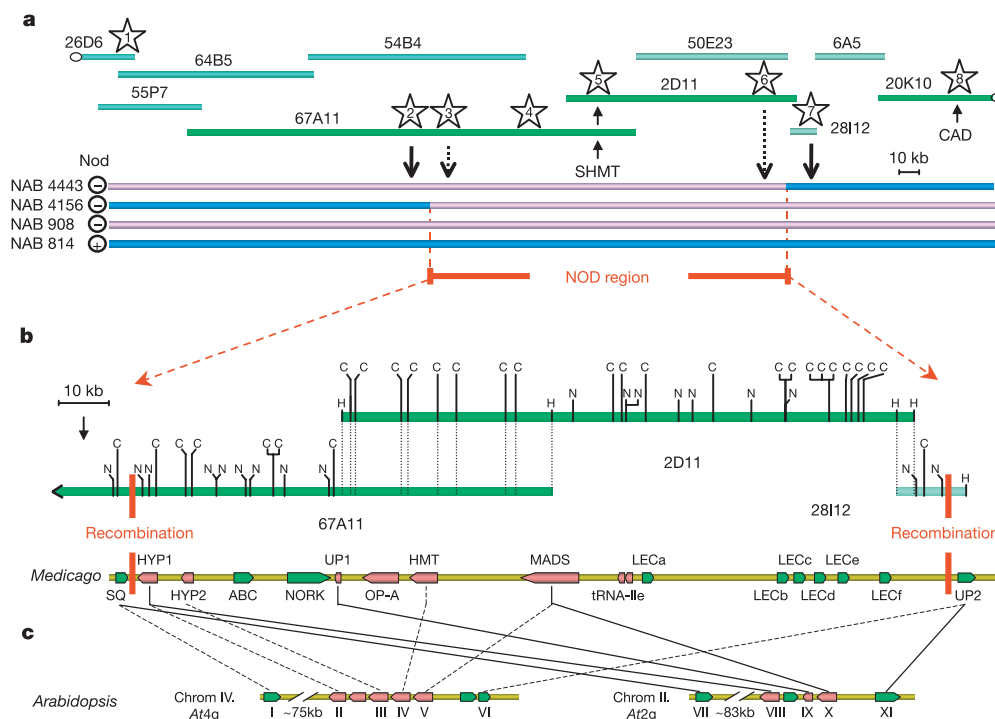


Figure 1 Genetic and physical map around the Nod locus. **a**, Primary BAC clones (67A11, 2D11 and 20K10) are coloured dark green, overlapping BAC clones (26D6, 55P7, 64B5, 54B4, 50E23, 28I12 and 6A5) are coloured light green. Fragments of the BAC clones (indicated as numbers in stars) were used for fine mapping. The Nod⁻ and one Nod⁺ genotypes are shown as pink and blue bars, respectively. **b**, Physical map and the gene content of the sequenced region. Restriction cut sites for *Clal* (C) and *NheI* (N), and the

HindIII (H) ends of the overlapping BACs are shown in the upper portion. Orientations of the open reading frames are indicated with green and red colours. **c**, Comparison of gene content with *A. thaliana* chromosomes 2 and 4 showing similarities to the sequenced region of *M. truncatula* I, At4g37760; II, At4g37900; III, At4g37920; IV, At4g37930; V, At4g37940; VI, At4g37960; VII, At2g22830; VIII, At2g22660; IX, At2g22640; X, At2g22630; XI, At2g22620.

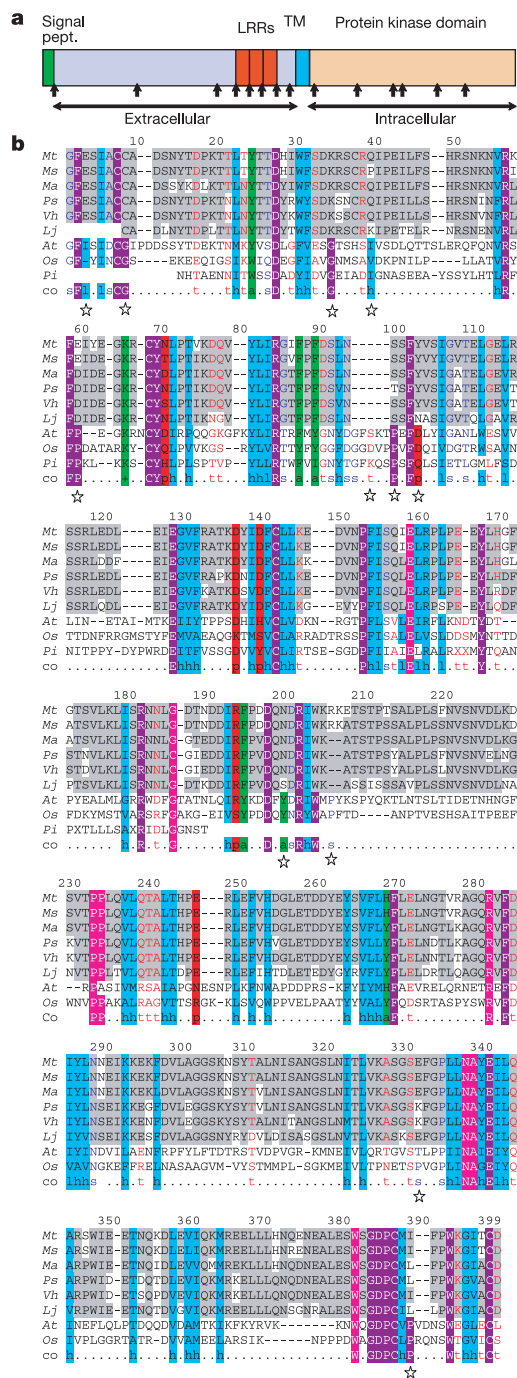


Figure 2 The NORK proteins. **a**, Schematic diagram of the NORK protein. The different domains are predicted on the basis of the deduced AA sequence. Vertical arrows indicate the location of the introns in the genomic DNA sequence. LRR, leucine-rich repeat; TM, transmembrane. **b**, Multiple alignment of the extracellular part of legume NORKs and three representatives of non-legume NSL proteins. *Mt*, *M. truncatula*; *Ms*, *M. sativa*; *Ma*, *Melilotus alba*; *Ps*, *Pisum sativum*; *Vh*, *Vicia hirsuta*; *Lj*, *Lotus japonicus*; *At*, *A. thaliana* LRRPK (accession number X97774); *Os*, *Oryza sativa* (accession number AAK92623); *Pi*, *Pinus taeda* (accession number BE758674); *co*, consensus. AA residues are coloured according to an 80% consensus (calculation included more NSL sequences, see Supplementary Information): hydrophobic (h: ACFGHIKLMRTWVY) and their aliphatic subset (l: ILV) with turquoise shading; aromatic (a: FHWY) and positive (+: HKR) with green shading; polar (p: CDEHKQNRST) with red shading; turn-like (t: ACDEGHKQNRST) in red; small (s: ACDGNPSTV) in blue. Dark violet background with white letters shows the position of 100% conserved AA, whereas pink background features the 80% consensus AAs. Grey background highlights conservation within legume NORKs, stars indicate the positions where AAs of the legume domain differ from the consensus of other NSL sequences.

2a, the predicted proteins possess an amino-terminal signal peptide motif, a putative extracellular domain containing three leucine-rich repeat (LRR) motifs¹⁴, as well as a transmembrane segment and an intracellular domain with typical serine/threonine protein kinase signatures^{15,16}. Several putative phosphorylation sites (see Supplementary Information) suggest a possible role also for (auto)phosphorylation in the protein's function. The predicted LRR motifs may indicate interaction with other protein(s)¹⁷. Although the N-terminal extracellular domain is well conserved among the legume proteins (Fig. 2b), this stretch, roughly 400 amino acids (AA) long, did not show homology to any known domain.

To further substantiate that in fact the mutation in the *NORK* gene was responsible for the non-nodulation phenotype, a similar sequencing strategy was undertaken on different *Nod*⁻ mutants. In *M. truncatula* and *P. sativum*, mutagenesis programmes have yielded non-nodulation mutants with phenotypes and syntenic map positions that are suggestive of orthology with the alfalfa *nn1* gene^{18,19}. Each sequenced *NORK* gene from the *Nod*⁻ alleles of *M. truncatula* and *P. sativum* contained single base-pair alterations that are predicted to abolish *NORK* function. The nucleotide changes and their possible effects are shown in Table 1. Among the seven mutant alleles of *NORK*, six resulted in a *Nod*⁻/*Myc*⁻ phenotype. It is interesting to note that a non-synonymous substitution (see Table 1) in the mutant *M. truncatula* R38 resulted in *Nod*⁻/*Myc*⁺ phenotype. This suggests a composite role for *NORK*, branching the signal towards nodule formation and mycorrhiza colonization. The identification of molecular lesions in seven alleles, obtained from independent non-nodulation mutants in three legume species, provides strong circumstantial evidence of a role for *NORK* in nodulation and mycorrhizal association.

To directly test this hypothesis, we analysed the ability of the wild-type *NORK* gene of *M. truncatula* to complement the non-nodulation mutant, TR25, following transformation by *Agrobacterium rhizogenes*²⁰. Roots appearing on the seedlings of the TR25 *Nod*⁻ mutants after *Agrobacterium* infection were tested for nodulation by inoculating with *S. meliloti*. Plants transformed by *A. rhizogenes* carrying an empty vector control tested positive for β -glucuronidase (GUS) staining of transgenic roots, but were unable to form nodules. By contrast, TR25 plants infected by *A. rhizogenes* carrying the *NORK* gene on the same plasmid tested positive for GUS staining and developed nodules containing bacteroids (see Supplementary Information). The combination of the positive complementation tests and the genetic evidence provided by sequencing of multiple independent non-nodulation mutants, suggest that *NORK* is an essential component of the host plant nodulation signalling pathway.

To determine the range of legume and non-legume species in which the *NORK* protein is conserved, we used the DNA region coding for the postulated extracellular domain of *NORK* as a hybridization probe on Southern blots of total genomic DNA. Specific hybridizing bands could be detected (Fig. 3) in each of the legume genera tested (*Sesbania*, *Cassia*, *Trifolium*, *Desmodium*, *Vicia*, *Melilotus*, *Vigna*, *Macroptilium*, *Lotus*, *Glycine*, *Pisum*, *Phaseolus*, *Medicago*). Homology in *Cassia emerginata* is particularly interesting because this Caesalpinoid legume species lacks symbiotic nodule formation²¹. Genetic mapping of the *NORK* gene in *Medicago* demonstrated only one segregating locus (data not shown). Similarly, in most other legume species that we tested, only one strong hybridizing band was detected, suggesting the presence of a single copy gene. In a few cases, however, more than one equally hybridizing band could be detected, indicating either a single copy gene with an internal restriction site, or the presence of a small *NORK* gene family. Representative members of non-legume plants (maize, tobacco, wheat, rice and *Arabidopsis*) did not show a specific hybridization signal, indicating that high homology to *NORK* is restricted primarily to legumes.

Despite the absence of high nucleotide similarity outside of

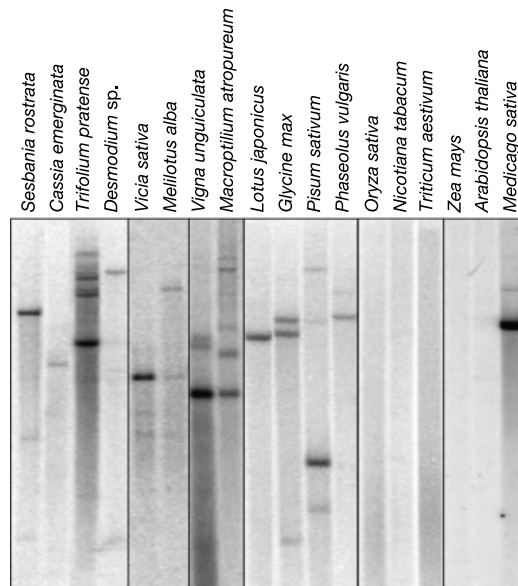


Figure 3 Genomic hybridization with the *NORK* gene to different legume and non-legume plants. Total DNA isolated from young leaf tissue of the plants highlighted were digested by *Eco*RI (except that *Cassia emerginata* DNA was digested by *Hind*III), at equivalent

amounts. A 1,700-bp PCR fragment containing only the putative extracellular part of the protein was used as the *NORK*-specific probe.

legumes revealed by DNA hybridization, BLASTX searches of the *NORK* unique extracellular domain identified many homologous sequences from other plant species at the protein level. We refer to these homologues as NSL (*NORK* extracellular sequence-like) sequences to distinguish them from other non-related receptor kinase homologues. The multiple alignment of these NSL sequences revealed a number of conserved AA residues throughout this domain (Fig. 2b). The presence of about 50 different genomic NSL sequences in the *Arabidopsis* genome indicates the existence of a large gene family containing the NSL motif. Almost all of these *Arabidopsis* NSL genes also have predicted LRR, transmembrane and intracellular kinase domains (NSL-RK). These receptor kinases were recently classified into the LRR-type category on the basis of their kinase sequences²². In addition to genomic sequences, many of the *Arabidopsis* homologues are represented as expressed sequences in GenBank, suggesting that they are likely to code for functional proteins. Besides information on tissue specific expression of two genes^{23,24}, no mutation/phenotype for any of these NSL sequences has been described. One full-length *Arabidopsis* complementary DNA sequence (AY059761) is predicted to code for a protein containing only the NSL domain. This example suggests that NSL

proteins lacking the LRR, transmembrane and intracellular kinase domains may also have biological function.

NSL genes were found (based on either genomic or expressed sequence tag (EST) sequences) not only in other dicotyledonous plants, but also in monocotyledonous plants such as rice, maize and wheat, and a homologue was also identified in Gymnosperms (see Fig. 2b and Supplementary Information). Owing to the limited number of sequences available from these organisms (in some cases only partial cDNA sequences were available), the number of NSL genes in these species could not be estimated. On the basis of BLASTX analysis, NSL homologues were not identified in animals, lower eukaryotes or prokaryotes, but we can not exclude the possibility that more highly diverged members of the NSL family might also be present in these non-plant genomes.

The mechanism(s) by which bacterial Nod factors are recognized by legume plants has remained obscure. If *NORK* proteins interact directly with Nod factors, then differences in the AA sequence of *NORK* extracellular domain from different legumes might reflect specificity for Nod-factor structure. It is also possible that Nod factors are bound by another protein, and that events occurring after ligand binding require the activity of *NORK*. In fact, inter-

Table 1 Legume *NORK* genes sequenced with description of the mutations

WT <i>NORK</i> sequences	Mutant <i>NORK</i> sequences	Mutant phenotype	Nature of mutation
<i>Medicago sativa</i>	MIN-1008 (<i>nn1</i>)	Nod [−] Myc [−]	Nonsense mutation, premature translational termination before the kinase active site
<i>Medicago truncatula</i> A17	TR25, TR26 (<i>dmi2</i> alleles of Mt A17)	Nod [−] Myc [−]	Frame shift due to one basepair deletion, premature translational termination in the extracellular domain
<i>Medicago truncatula</i> A17	P1 (<i>dmi2</i> allele of Mt A17)	Nod [−] Myc [−]	G to A transition in the splice consensus site of intron 5 caused the loss of exon 6 from the mRNA
<i>Medicago truncatula</i> A17	R38* (<i>dmi2</i> allele of Mt A17)	Nod [−] Myc ⁺	Conserved Gly changed to Glu in the kinase subdomain X which might interact with the substrate ¹⁶
<i>Pisum sativum</i> cv. Frisson	P4 (<i>sym19</i> allele of <i>Ps</i> cv. Frisson)	Nod [−] Myc [−]	Strictly conserved Gly residue of the ATP-binding domain ¹⁶ changed to Glu
<i>Pisum sativum</i> cv. Frisson	P55 (<i>sym19</i> allele of <i>Ps</i> cv. Frisson)	Nod [−] Myc [−]	Asp of protein kinase subdomain VII ¹⁶ —involved in the binding of the ATP-chelating metal—changed to Asn

*N. Krishnamurthy, G. Endre, R. V. Penmetsa, D. R. Cook & K. VandenBosch. A. manuscript in preparation.

action of NORK with other proteins is suggested by the presence of LRR motifs and thus the extracellular domain may be involved in protein–protein interaction. On the basis of pharmacological studies, a G-protein coupled receptor-like protein has been suggested as a candidate for the Nod-factor receptor²⁵. Other candidate gene(s) for the Nod-factor receptor(s) include *DMI1* from *M. truncatula*^{18,26}, *SYM1* or *SYM5* from *L. japonicus*²⁷, and *SYM8* or *SYM10* from *P. sativum*²⁸, on the basis of their mutant phenotype. The map-based cloning of these genes (which is in progress) and their analysis will further elucidate the molecular bases of Nod-factor perception. The current study suggests that NORK is an integral component of the signalling pathway, in which the coordinated interaction of the components is needed for Nod-factor signal perception and transduction.

The NSL gene family—identified with the help of NORK extracellular domain—is broadly distributed in the plant kingdom. The AA conservation detected in this protein family suggests a similar molecular mechanism in their function. Thus NORK homologues may function in the perception and transduction systems for extracellular ligands in plants. We suggest that during the evolution of symbiosis an ancient NSL system was recruited by an external ligand, the bacterial Nod factors, leading to the developmental regulation of a new plant organ, the symbiotic nodule. □

Methods

Plant materials

M. sativa plants used for studies are individuals of a tetraploid mapping population described earlier¹². Wild-type and mutant seeds of *M. truncatula* A17 and *P. sativum* cv. Frisson were obtained from D. Cook (*M. truncatula* A17), J. Denarie (*M. truncatula* P1), G. Duc (*M. truncatula* TR25 and TR26, *P. sativum* cv Frisson, mutants P4 and P55) and K. VandenBosch (*M. truncatula* R38). Plants were grown as described previously¹².

Nucleic acid isolation and analysis

Genomic DNA isolation, restriction enzyme digestion, Southern blot, DNA hybridization (RFLP analysis) and polymerase chain reaction (PCR) amplification were carried out as described previously¹². BAC clone identification by hybridization and BAC DNA isolation were performed according to original descriptions¹³. BAC clones were subcloned after fragmentation by restriction enzyme digestion or random fragmentation by sonication into pUC19. Total RNA from roots was extracted using the High Pure RNA isolation kit (Roche). First strand cDNA was synthesized with the help of the RevertAid H Minus First Strand cDNA Synthesis Kit of MBI Fermentas. PCR amplifications of the overlapping fragments of the NORK gene from genomic and cDNA templates were set up in triplicate and each PCR fragment was cloned independently into pGEM-T-Easy vector (Promega). DNA sequencing was carried out on an ABI377 automatic sequencer using fluorescent dye terminators.

A. rhizogenes transformation

The *A. rhizogenes* hairy root transformation system recently developed for *M. truncatula* was used²⁰. The wild-type *M. truncatula* NORK gene with its own promoter sequence (more than 700 base pairs (bp) upstream from the proposed translation start site) was cloned on an 8.5-kb *NheI* fragment into the modified pPR97 plant transformation vector²⁹ carrying a constitutively expressed *uidA* gene (clone pBRC1720). *A. rhizogenes* carrying the pBRC1720 clone or empty pPR97 vector (negative control) were used to transform seedlings of the *M. truncatula* TR25 Nod[−] mutants. Experiments were repeated twice. In total, 25 GUS-positive roots developed on 10 treated plants, on which 20 nodules appeared that contained bacteroids. In control plants GUS-positive roots also developed but no nodule formation was detected.

Computational analyses

BLAST searches were conducted against the non-redundant and EST databases at NCBI (www.ncbi.nlm.nih.gov/BLAST). Multiple alignment of the NSL sequences was constructed using the ClustalW program³⁰. Protein domain predictions were carried out at www.smart.embl-heidelberg.de and www.ch.embnet.org. Consensus sequence was calculated from the multiple alignment at www.bork.embl-heidelberg.de/Alignment/consensus.html.

Received 6 February; accepted 23 April 2002; doi:10.1038/nature00842.

1. Lerouge, P. *et al.* Symbiotic host-specificity of *Rhizobium meliloti* is determined by a sulphated acylated glucosamine oligosaccharide signal. *Nature* **344**, 781–784 (1990).
2. Schultze, M. & Kondorosi, A. Regulation of symbiotic root nodule development. *Annu. Rev. Genet.* **32**, 33–57 (1998).
3. Schultze, M. & Kondorosi, A. What makes nodulation signals host-plant specific? *Trends Microbiol.* **3**,

- 370–372 (1995).
4. Peterson, M. A. & Barnes, D. K. Inheritance of ineffective nodulation and non-nodulation traits in alfalfa. *Crop Sci.* **21**, 611–616 (1981).
5. Dudley, M. E. & Long, S. R. A non-nodulating alfalfa mutant displays neither root hair curling nor early cell division in response to *Rhizobium meliloti*. *Plant Cell* **1**, 65–72 (1989).
6. Ehrhardt, D. W., Wais, R. & Long, S. R. Calcium spiking in plant root hairs responding to *Rhizobium* nodulation signals. *Cell* **85**, 673–681 (1996).
7. Felle, H. H., Kondorosi, É., Kondorosi, Á. & Schultze, M. Rapid alkalization in alfalfa root hairs in response to rhizobial lipochitoooligosaccharide signals. *Plant J.* **10**, 295–301 (1996).
8. Truchet, G. *et al.* Alfalfa nodulation in the absence of *Rhizobium*. *Mol. Gen. Genet.* **219**, 65–68 (1989).
9. Caetano-Anollés, G., Joshi, P. A. & Gresshoff, P. M. in *New Horizons in Nitrogen Fixation* (eds Palacios, R., Mora, J. & Newton, W. E.) 297–302 (Kluwer Academic, Dordrecht, 1993).
10. Bradbury, S. M., Peterson, R. L. & Bowley, S. R. Interaction between three alfalfa nodulation genotypes and two *Glomus* species. *New Phytol.* **119**, 115–120 (1991).
11. Gianinazzi-Pearson, V. Plant cell responses to arbuscular mycorrhizal fungi—getting to the roots of the symbiosis. *Plant Cell* **8**, 1871–1883 (1996).
12. Endre, G. *et al.* Genetic mapping of the non-nodulation phenotype of the mutant MN-1008 in tetraploid alfalfa (*Medicago sativa*). *Mol. Genet. Genomics* **266**, 1012–1019; advance online publication, 23 January (2002) (doi:10.1007/s00438-001-0628-3).
13. Nam, Y. W. *et al.* Construction of a bacterial artificial chromosome library of *Medicago truncatula* and identification of clones containing ethylene response genes. *Theor. Appl. Genet.* **98**, 638–646 (1999).
14. Kobe, B. & Deisenhofer, J. The leucine-rich repeat: a versatile binding motif. *Trends Bio. Sci.* **19**, 415–421 (1994).
15. Hanks, S. K. & Quinn, A. M. Protein kinase catalytic domain sequence database: identification of conserved features of primary structure and classification of family members. *Methods Enzymol.* **200**, 38–61 (1991).
16. Schenk, P. W. & Snaar-Jagalska, B. E. Signal perception and transduction: role of protein kinases. *Biochem. Biophys. Acta* **1449**, 1–24 (1999).
17. Hurlley, S. M. & Helenius, A. Protein oligomerization in the endoplasmic reticulum. *Annu. Rev. Cell Biol.* **5**, 277–307 (1989).
18. Catoira, R. *et al.* Four genes of *Medicago truncatula* controlling components of a Nod factor transduction pathway. *Plant Cell* **12**, 1647–1665 (2000).
19. Schneider, A. *et al.* Genetic mapping and functional analysis of a nodulation-defective mutant (*sym19*) of pea (*Pisum sativum* L.). *Mol. Gen. Genet.* **262**, 1–11 (1999).
20. Boisson-Dernier, A. *et al.* *Agrobacterium rhizogenes*-transformed roots of *Medicago truncatula* for the study of nitrogen-fixing and endomycorrhizal symbiotic associations. *Mol. Plant-Microbe Interact.* **14**, 695–700 (2001).
21. Doyle, J. J. Phylogenetic perspectives on nodulation: evolving views of plants and symbiotic bacteria. *Trends Plant Sci.* **3**, 473–478 (1998).
22. Shiu, S.-H. & Blecker, A. B. Receptor-like kinases from *Arabidopsis* form a monophyletic gene family related to animal receptor kinases. *Proc. Natl Acad. Sci. USA* **98**, 10763–10768 (2001).
23. Hajoui, T., Michels, R. & Gepstein, S. Cloning and characterization of a receptor-like protein kinase gene associated with senescence. *Plant Physiol.* **124**, 1305–1314 (2000).
24. Deeken, R. & Kaldenhoff, R. Light-repressible receptor protein kinase: a novel photo-regulated gene from *Arabidopsis thaliana*. *Planta* **202**, 479–486 (1997).
25. Pingret, J. L., Journet, E. P. & Barker, D. G. *Rhizobium* Nod factor signalling. Evidence for a G protein-mediated transduction mechanism. *Plant Cell* **10**, 659–672 (1998).
26. Wais, R. J. *et al.* Genetic analysis of calcium spiking responses in nodulation mutants of *Medicago truncatula*. *Proc. Natl Acad. Sci. USA* **97**, 13407–13412 (2000).
27. Stougaard, J. Genetics and genomics of root symbiosis. *Curr. Opin. Plant Biol.* **4**, 328–335 (2001).
28. Walker, S. A. *et al.* Dissection of nodulation signalling using pea mutants defective for calcium spiking induced by Nod factors and chitin oligomers. *Proc. Natl Acad. Sci. USA* **97**, 13413–13418 (2000).
29. Szabados, L., Charrier, B., Kondorosi, A., Debruijn, F. J. & Ratet, P. New plant promoter and enhancer testing vectors. *Mol. Breeding* **1**, 419–423 (1995).
30. Thompson, J. D., Higgins, D. G. & Gibson, T. J. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucl. Acids Res.* **22**, 4673–4680 (1994).

Supplementary Information accompanies the paper on Nature's website (<http://www.nature.com/nature>).

Acknowledgements

We thank D. Cook for comments, suggestions and help in finalizing the manuscript, A. Kondorosi for critical reading of the manuscript, A. Perhald for help in transformation experiments, and P. Kiss, S. Jenei, K. Lehoczy, Z. Liptay, P. Somkúti and A. Lengyel for technical assistance. This work was supported by the BRC, Szeged, Hungary, the Bátyai-Holzer Foundation, the Hungarian Academy of Sciences, the Hungarian Scientific Research Fund, the Hungarian National Committee for Technical Development, the Szechenyi Fund of the Hungarian Ministry of Education, the EuDicoMap and the Medicago projects of the European Union.

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to G.B.K. (e-mail: kgb@nucleus.szbk.u-szeged.hu or/and e-mail: gbkiss@ucdavis.edu). Accession numbers in the EMBL/GenBank/DBJ databases for supporting data sets are as follows: AJ418368; AJ418377; AJ418369; AJ418370; AJ418371; AJ418372; AJ418373; AJ418374; AJ418375; AJ418376; AJ418367; AJ428990; AJ428991.

A large nucleolar U3 ribonucleoprotein required for 18S ribosomal RNA biogenesis

François Dragon^{*†}, Jennifer E. G. Gallagher[‡],
 Patricia A. Compagnone-Post^{*}, Brianna M. Mitchell^{*},
 Kara A. Porwancher[‡], Karen A. Wehner[‡], Steven Wormsley[§],
 Robert E. Settlege^{||}, Jeffrey Shabanowitz[¶], Yvonne Osheim[#],
 Ann L. Beyer[#], Donald F. Hunt[¶] & Susan J. Baserga^{*‡}

^{*} Department of Therapeutic Radiology, [‡] Department of Genetics, and
[§] Department of Cell Biology, Yale University School of Medicine, New Haven,
 Connecticut 06520-8040, USA
^{||} ProteoMS, LLC, Charlottesville, Virginia 22903, USA
[¶] Department of Chemistry, [#] Department of Microbiology, and [☆] Department of
 Pathology, University of Virginia, Charlottesville, Virginia 22904, USA

Although the U3 small nucleolar RNA (snoRNA), a member of the box C/D class of snoRNAs, was identified with the spliceosomal small nuclear RNAs (snRNAs) over 30 years ago^{1,2}, its function and its associated protein components have remained more elusive. The U3 snoRNA is ubiquitous in eukaryotes and is required for nucleolar processing of pre-18S ribosomal RNA in all organisms where it has been tested^{3,4}. Biochemical and genetic analyses suggest that U3-pre-rRNA base-pairing interactions mediate endonucleolytic pre-rRNA cleavages⁵. Here we have purified a large ribonucleoprotein (RNP) complex from *Saccharomyces cerevisiae* that contains the U3 snoRNA and 28 proteins. Seventeen new proteins (Utp1–17) and Rrp5 were present, as were ten known components. The Utp proteins are nucleolar and specifically associated with the U3 snoRNA. Depletion of the Utp proteins impedes production of the 18S rRNA, indicating that they are part of the active pre-rRNA processing complex. On the basis of its large size (80S; calculated relative molecular mass of at least 2,200,000) and function, this complex may correspond to the terminal knobs present at the 5' ends of nascent pre-rRNAs. We have termed this large RNP the small subunit (SSU) processome.

Previous U3 snoRNP purification strategies have used RNA affinity approaches and have resulted in the identification of only a small number of components^{5,6}. We chose a protein-based affinity purification method that relied on simultaneous epitope tagging of two components: Nop5/58, which is common to the box C/D snoRNPs (TAP tag⁷), and the U3-specific protein Mpp10 (Flag tag). The first step immobilizes the box C/D snoRNPs through the protein A portion of the TAP tag. The snoRNPs are then released by TEV protease cleavage, maintaining native conditions. To isolate the U3 snoRNP, the box C/D snoRNPs are then applied to an anti-Flag antibody column and released with an excess of Flag peptide. The RNA composition of the resulting fractions was analysed (Fig. 1). The predominant enriched small RNA present in the final eluted material is the U3 snoRNA (lane 4), whereas the box C/D snoRNAs are enriched only after the first step (lane 3). The 5S and 5.8S rRNAs are also present in the eluate, although they are not enriched with respect to their relative abundance in the starting extract (lane 2).

U3 snoRNA-associated protein components in the Flag affinity column eluate were digested with trypsin and the resulting peptides were analysed by nanoflow high-performance liquid chromatography/electrospray ionization mass spectrometry. Twenty-eight proteins ranging in relative molecular mass (M_r) from 13,000 to 200,000 were identified (Supplementary Information Table 1). All but one are coded for by essential yeast genes. Ten proteins had been

previously described as U3 snoRNA-associated: four proteins common to the box C/D snoRNPs (Nop1, Nop56, Nop5/58 and Snu13) and six proteins specific to the U3 snoRNP (Sof1, Mpp10, Imp3, Imp4, Dhr1, Rrp9). We did not find three proteins previously demonstrated to be U3-associated: Lcp5 (ref. 8), Rcl1 (ref. 9) and Bms1 (ref. 10); however, this does not rule out their presence in the complex as they may be transiently associated, present in substoichiometric amounts or lost during the purification. One protein, Rrp5, had previously been shown to be required for pre-18S processing. Seventeen proteins (U three protein, Utp1–17) were found that had not previously been shown to be associated with U3 or implicated in pre-rRNA processing events. In addition, our analysis yielded five small subunit ribosomal proteins (Rps4, Rps6, Rps7, Rps14 and Rps28). Three proteins (YER087w, YPL110c, YMR029c) subsequently proved to be contaminants, as tagged versions did not co-immunoprecipitate Mpp10 and because the two that could be immunolocalized were found to be cytoplasmic (YER087w and YPL110c).

Bioinformatics analysis of the new protein components revealed motifs characteristic of an RNA-protein machine (Supplementary Information Table 1). Fourteen of the 17 Utp proteins bear protein-protein interaction domains (WD repeats, coiled-coil domains, HEAT repeats and a crooked-neck-like (crn-like) tetratricopeptide repeat (TPR)). The crn-like TPR is found in several proteins involved in other RNA processing events such as pre-mRNA splicing (Prp42, Prp6 and Clf1) and polyadenylation (RNA14)¹¹. Two components have motifs characteristic of catalytically active proteins, an adenylate binding site and an ATP/GTP-binding site (P loop), although none so far identified bears a motif characteristic of an RNA endonuclease. Two of the Utp proteins have been implicated in other aspects of cellular metabolism: Utp17 was co-purified with the RENT complex, which is essential for exit from mitosis¹², and Utp3 was discovered because it disrupts transcrip-

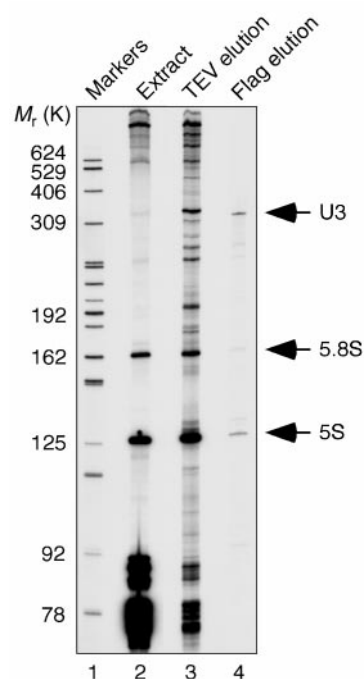


Figure 1 RNA composition of fractions from the purification of the U3 snoRNP. RNA was extracted from the starting material (extract, lane 2), from the first purification step that enriches for the box C/D snoRNAs (TEV elution, lane 3), and from the final eluate (Flag elution, lane 4). The RNA was directly labelled by ³²p-labelled pCp and T4 RNA ligase, and analysed on an 8% denaturing polyacrylamide gel.

[†] Present address: Université du Québec à Montréal, Montréal H3B 3P8, Canada

tional silencing on overexpression¹³. For 11 of the 17 Utp proteins, probable human homologues exist; two of the remaining six Utp proteins have invertebrate homologues. For one of these, a probable homologue exists in *Schizosaccharomyces pombe*. This degree of evolutionary conservation suggests preservation of the fundamental mechanism of pre-rRNA processing from unicellular organisms to humans.

The previously identified U3 snoRNP-specific proteins are nucleolar and co-immunoprecipitate the U3 snoRNA. To verify that the new Utp proteins meet these criteria, we carried out the following experiments. The 17 Utp proteins were tagged with the haemagglutinin (HA)-epitope by chromosomal integration. Their subcellular localization was evaluated by indirect immunofluorescence microscopy. All 17 Utp proteins co-localized with the Mpp10 protein, indicating their nucleolar localization (Fig. 2a). Consistent with the co-localization studies, all 17 Utp proteins, as well as Rrp5, co-immunoprecipitated the Mpp10 protein (Fig. 2b). Other non-U3-associated nucleolar proteins, such as Gar1 and Nop7, did not co-immunoprecipitate Mpp10, suggesting that the Utp proteins and Rrp5 belong to a distinct nucleolar complex. Finally, all 17 Utp proteins and Rrp5 co-immunoprecipitated the U3 snoRNA (Fig. 2c). Furthermore, the protein–RNA association was stable after washes in 400 mM NaCl. In contrast, there was essentially no co-immunoprecipitation of the U14, snR10 or snR30 snoRNAs with any of the Utp proteins or with Rrp5 at either of the salt concentrations used (data not shown). The Utp proteins and Rrp5 are therefore part of a nucleolar complex that contains both Mpp10 and the U3 snoRNA but excludes many other nucleolar factors.

To define the function of the Utp proteins we constructed yeast

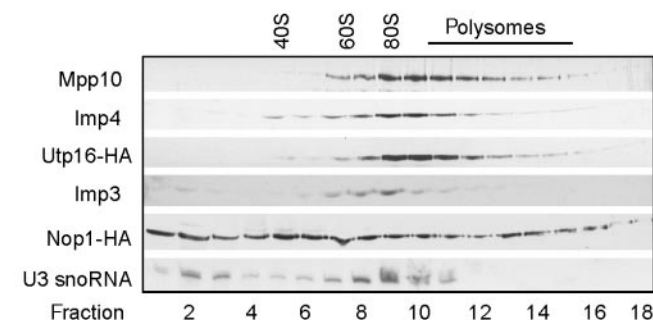


Figure 3 The SSU processome sediments at 80S on sucrose gradients. Yeast extracts were analysed on 10–47% sucrose gradients. Anti-Mpp10, anti-Imp3, anti-Imp4 and anti-HA antibodies were used to detect the indicated proteins in each fraction by western blotting. Northern blotting was used to detect the U3 snoRNA. The migration of the 40S, 60S and 80S ribosomal subunits is specified. Utp16 and Nop1 proteins were triply tagged by HA.

strains in which each of the essential proteins could be conditionally depleted by growth in glucose. After depletion, we assessed the levels of 25S and 18S rRNAs by northern blotting. Depletion of each of the essential Utp proteins resulted in reduction of 18S rRNA levels with respect to the levels of 25S rRNA (Fig. 2d). Depletion of each Utp protein did not, however, affect the levels of the U3 snoRNA (Supplementary Information Fig. 1). These results suggest that the reduction in 18S rRNA levels upon Utp protein depletion is not due to a general defect in U3 snoRNP biogenesis but instead

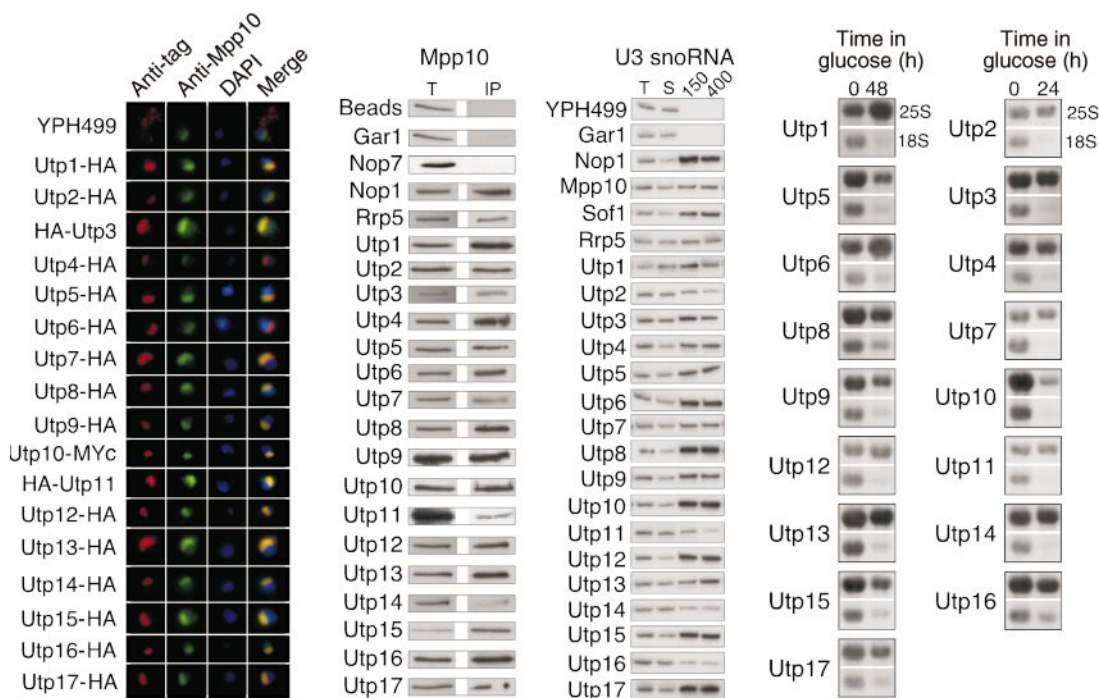


Figure 2 Function of the new components of the SSU processome. **a**, The Utp proteins are nucleolar. Anti-HA or anti-Myc antibodies were used to detect the tagged Utp proteins (red), whereas anti-Mpp10 antibodies were used to decorate the nucleolus (green). DAPI was used to stain the nucleus (blue). Utp proteins were triply tagged by HA or Myc. **b**, The Utp proteins and Rrp5 are complexed with Mpp10. Anti-HA immunoprecipitations performed on cell extracts were analysed for the presence of Mpp10 by western blotting with anti-Mpp10 antibodies. T, total (5% of the input for immunoprecipitation); IP, immunoprecipitate. **c**, The Utp proteins and Rrp5 are associated with U3 snoRNA. Anti-HA

immunoprecipitations were performed on cell extracts and were washed at 150 mM (150) or 400 mM (400) NaCl. RNA was extracted and analysed for the presence of the U3 snoRNA by northern blotting. T, total (20% of the input); S, supernatant (20%). **d**, The Utp proteins are required for 18S rRNA biogenesis. Yeast strains conditionally expressing Utp proteins were grown in medium containing galactose/raffinose (0 h), but switched to glucose for protein depletion (24 or 48 h). RNA was extracted at the indicated time points and analysed for the presence of 25S and 18S rRNAs by northern blot.

is due to a specific protein deficiency. The Utp proteins are therefore required for biogenesis of the small ribosomal subunit RNA, as are all of the other known U3 snoRNP-specific components.

The finding of 28 U3-associated proteins prompted us to examine the size of the U3 snoRNA-associated protein complex on sucrose gradients (Fig. 3). The U3-specific proteins Mpp10, Imp4, Imp3, Utp16 (as an example of a new protein) and the U3 snoRNA co-migrate at a size that approximately corresponds to that of 80S ribosomes. In contrast, Nop1 (fibrillarin), which is a protein common to all box C/D snoRNPs, does not share the same sedimentation profile. Therefore the large U3-protein complex is approximately the same size as 80S ribosomes.

Our results, combined with those of others, demonstrate that the 80S U3-protein complex has at least 28 physically and functionally associated proteins. Thus, the active U3 particle is not a small nucleolar RNP but rather a large nucleolar RNP. We have termed this complex the small subunit (SSU) processome because its components are required for pre-18S rRNA processing^{14,15}. By analogy with the spliceosome, the SSU processome may be present in forms that differ slightly in composition, and it may be a dynamic entity. Similarly, there may be other proteins and snoRNPs transiently associated with it or present in the SSU processome in substoichiometric amounts.

The presence of terminal knobs on amphibian pre-rRNA transcripts was first noted in 1969 (ref. 16), and the structures were termed 'Christmas trees' for their characteristic appearance. More than 20 years later, it was proposed that the terminal knobs included some of the pre-rRNA processing machinery¹⁷. Electron micrographs of chromatin spreads from yeast indicate that their nascent rRNA transcripts, like those in amphibians, also possess 5' terminal knobs (Fig. 4a). In a strain where either the U3 snoRNA or Utp7 can be conditionally depleted by growth in glucose, we found that their depletion leads to loss of the terminal knobs on nascent pre-rRNAs (Fig. 4b, c). Depletion of Imp3 and Imp4 also leads to loss of the terminal knobs (data not shown). Switch of a parent strain from galactose to glucose, however, has no effect on the terminal knobs (data not shown); neither does a mutation in Mpp10 that does not lower protein levels (data not shown)¹⁸. Therefore the presence both of the U3 snoRNA and of SSU processome proteins is required for

terminal knob formation.

The size of the yeast terminal knobs measured using electron micrographs is variable, with an average size (42×34 nm) similar to that of ribosomes measured in the same way (28×30 nm). The size of the SSU processome estimated by sedimentation analysis (80S) and its calculated minimal mass (M_r 2,200,000, assuming each component is present once) are also similar to that of ribosomes. Therefore the SSU processome is roughly the same size as the terminal knobs. Owing to the approximate size correlation between the SSU processome and the terminal knobs, the requirement for the U3 snoRNA, Utp7, Imp3 and Imp4 in terminal knob formation, and the requirement for the SSU processome and the terminal knob in rRNA biogenesis¹⁷, we propose that the terminal knobs on nascent pre-rRNAs may represent the SSU processome. Further support comes from the observation that one of the SSU processome components, Imp4, co-immunoprecipitates the nascent 35S rRNA near the 5' end¹⁹, as might be expected if it were a component of the terminal knob.

Why are there so many proteins required for the pre-rRNA cleavage reactions? The results presented here indicate that there are at least 28 proteins in the SSU processome. All of the proteins essential for growth are required for 18S rRNA biogenesis. We suggest that the complexity of the SSU processome reflects multiple functions. For example, the SSU processome may also have a critical role in folding the pre-18S rRNA. Consistent with this is the proposal that specific U3 snoRNA sequences are required for formation of the central, conserved pseudoknot in 18S rRNA³. The SSU processome may therefore function both in pre-rRNA cleavage and as a pre-rRNA chaperone.

Other pre-ribosomal complexes have recently been isolated using protein affinity techniques coupled with mass spectrometry²⁰. In contrast to the function of these complexes in later steps in pre-rRNA processing or in transport, the SSU processome functions in the earliest steps in ribosome biogenesis. Of note, none of the non-ribosomal protein components of the SSU processome were found in the other identified complexes. These distinctions suggest that the different pre-ribosomal particles are each tailored for a specific function.

This work identifies the SSU processome as the third large RNP in

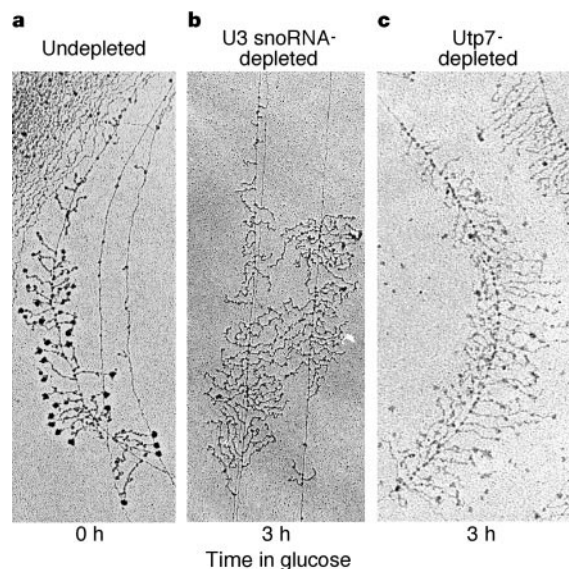


Figure 4 The U3 snoRNA and Utp7 are required for terminal knob formation on nascent pre-rRNAs. Yeast strains conditionally expressing either the U3 snoRNA or Utp7 from a galactose promoter were used to make the chromatin spreads. Strains were undepleted (a) or depleted for U3 snoRNA (b) or Utp7 (c). For depletion, the strains were switched from

growth in galactose to growth in glucose. Chromatin spreads were made before (0 h) and after (3 h) the switch to glucose and analysed by electron microscopy. Scale, the width of panel a is 0.85 μ m.

eukaryotic cells, the two others being the ribosome and the spliceosome. Our results show that the SSU processome is an RNP larger and more complex than originally thought, with a definitive role in pre-rRNA processing and a putative role in pre-rRNA folding. Thus, it takes a large RNP (the SSU processome) to make a large RNP (the ribosome). □

Methods

For purification of the U3 snoRNP, mass spectrometry and bioinformatics analysis, see Supplementary Information.

Validation and functional analysis

Saccharomyces cerevisiae strains expressing carboxy-terminal triple-HA-tagged versions of the Utp proteins, Gar1, Nop7, Nop1, Sof1 and Rrp5 were created by homologous recombination of polymerase chain reaction products at chromosomal loci in strain YPH499 (ref. 21). Strains expressing amino-terminal triple-HA-tagged versions of the Utp proteins were similarly created²². The genes N-terminally tagged were expressed under the control of galactose promoters to generate strains conditionally expressing the Utp proteins. In both cases, *kanMX6* was used as the selection marker. Utp10 in Fig. 2a was tagged at its C terminus by chromosomal integration of triple-Myc using *URA3* as a marker²³. Successful tagging was verified by western blot analysis. The HA-tagged Mpp10 was obtained from the laboratory of M. Snyder²⁴.

Yeast expressing tagged proteins were processed for indirect immunofluorescence microscopy as described²⁵. Mouse anti-HA monoclonal antibody HA.11 (diluted 1:1,000; Covance) and rabbit anti-Mpp10 polyclonal antibodies²⁶ (diluted 1:2,000) were detected with tetramethyl rhodamine isothiocyanate-conjugated goat anti-mouse immunoglobulin- γ (IgG) (diluted 1:100) and fluorescein isothiocyanate-conjugated donkey anti-rabbit IgG (diluted 1:200) secondary antibodies (Jackson ImmunoResearch). For localization of tagged Utp10, mouse anti-Myc monoclonal antibody 9E10 was used in place of the anti-HA antibody. Nuclear DNA was stained with 4,6-diamidino-2-phenylindole (DAPI, $1 \mu\text{g ml}^{-1}$; Sigma).

For analysis of Mpp10 co-immunoprecipitation, anti-HA (12CA5) immunoprecipitations were carried out on extracts from strains bearing tagged proteins and western blotted with anti-Mpp10 antibodies as described²⁷. Except for Utp11, each immunoprecipitation was performed with strains expressing C-terminal triple-HA-tagged proteins. Utp11 was triply HA-tagged at the N terminus.

For analysis of snoRNA co-immunoprecipitation, extracts were made and immunoprecipitations performed as described²⁷, except that HA beads were prepared by incubating 200 μl anti-HA (12CA5) with 3 mg protein A Sepharose beads (Amersham Pharmacia Biotech) for 1–2 h at room temperature. The HA-tagged strains are described above.

Depletion was carried out on strains expressing Utp proteins under the control of galactose promoters. The strains were grown in medium containing galactose and raffinose, and then switched to glucose for depletion. For some strains, growth was severely impeded after 24 h in glucose, and RNA was extracted. For others, no change in growth was observed after 24 h in glucose, and the yeast were re-diluted and grown for another 24 h (48 h total) before RNA extraction. For these strains, growth was severely impeded after dilution at 24 h of growth in glucose. All strains were grown at 30 °C except for the strain expressing Utp16 from a galactose promoter. This strain is cold-sensitive when grown in glucose and was grown at room temperature. We carried out RNA extraction and northern blots as described²⁷.

Sucrose gradients

Extracts were prepared²⁷ and analysed on 10–47% sucrose gradients in 25 mM Tris pH 7.6, 150 mM KCl, 10 mM MgCl_2 , 0.1% NP-40. The gradients were spun in a SW41 rotor at 39,000 r.p.m. (260,000g) for 2 h. We collected 18 570- μl fractions. The migration of the 40S, 60S and 80S ribosomes was determined from the ultraviolet profile provided by the ISCO Model 185 density gradient fractionator. Fractions were analysed for the indicated proteins by western blotting with rabbit anti-Mpp10 (ref. 26), guinea-pig anti-Imp3, guinea-pig anti-Imp4 or anti-HA (12CA5). The guinea-pig anti-Imp3 and Imp4 antibodies were raised to recombinant proteins expressed and purified from *Escherichia coli* as described²⁶. The U3 snoRNA was detected by northern blotting²⁷.

Chromatin spreads and electron microscopy

We used strain JH84 for U3 snoRNA depletion^{28,29}. In this strain the *U3A* snoRNA gene is under the control of a galactose promoter and the *U3B* gene is disrupted. The strain for depletion of Utp7 was constructed as described above. Depletion of the U3 snoRNA and Utp7 was performed by growth in glucose for 3 h. For chromatin spreads, yeast cultures were grown in YP medium plus galactose or glucose plus 1 M sorbitol to an absorbance at 600 nm of about 0.4. One millilitre of culture was digested with 5 mg zymolyase for 4 min at 30 °C. The yeast were then pelleted and 1 ml of 0.025% TritonX-100 at pH 9 was added to the pellet. After resuspension the yeast solution was mixed with 3 ml of 0.025% Triton and allowed to disperse for 20 min with swirling. One-tenth volume of 0.1 M sucrose-10% formalin, pH 8.5, was then added and grids were made in the usual manner³⁰. Measurements of the size of the terminal balls were made on chromatin spreads of the YPH499 strain.

Received 20 December 2001; accepted 21 May 2002; doi:10.1038/nature00769.
Published online 9 June 2002.

1. Weinberg, R. A. & Penman, S. Small molecular weight monodisperse nuclear RNA. *J. Mol. Biol.* **38**, 289–304 (1968).

- Nakamura, T., Prestayko, A. W. & Busch, H. Studies on nucleolar 4 to 6S ribonucleic acid of Novikoff hepatoma cells. *J. Biol. Chem.* **243**, 1368–1375 (1968).
- Venema, J. & Tollervey, D. Ribosome synthesis in *Saccharomyces cerevisiae*. *Annu. Rev. Genet.* **33**, 261–311 (1999).
- Maxwell, E. S. & Fournier, M. J. The small nucleolar RNAs. *Annu. Rev. Biochem.* **64**, 897–934 (1995).
- Watkins, N. J. *et al.* A common core RNP structure shared between the small nucleolar box C/D RNPs and the spliceosomal U4 snRNP. *Cell* **103**, 457–466 (2000).
- Lubben, B., Marshall, C., Rottman, N. & Luhrmann, R. Isolation of U3 snoRNP from CHO cells: a novel 55 kDa protein binds to the central part of U3 snoRNA. *Nucleic Acids Res.* **21**, 5377–5385 (1993).
- Rigaut, G., Shevchenko, A., Rutz, B., Wilm, M. & Seraphin, B. A generic protein purification method for protein complex characterization and proteome exploration. *Nature Biotechnol.* **17**, 1030–1032 (1999).
- Wiederkehr, T., Pretot, R. F. & Minvielle-Sebastia, L. Synthetic lethal interactions with conditional poly(A) polymerase alleles identify *LCP5*, a gene involved in 18S rRNA maturation. *RNA* **4**, 1357–1372 (1998).
- Billy, E., Wegierski, T., Nasr, F. & Filipowicz, W. Rcl1p, the yeast protein similar to the RNA 3'-phosphate cyclase, associates with U3 snoRNP and is required for 18S rRNA biogenesis. *EMBO J.* **19**, 2115–2126 (2000).
- Wegierski, T., Billy, E., Nasr, F. & Filipowicz, W. Bms1p, a G-domain-containing protein, associates with Rcl1p and is required for 18S rRNA biogenesis in yeast. *RNA* **7**, 1254–1267 (2001).
- Chung, S., McLean, M. R. & Rymond, B. C. Yeast ortholog of the *Drosophila* crooked neck protein promotes spliceosome assembly through stable U4/U6.U5 snRNP addition. *RNA* **5**, 1042–1054 (1999).
- Shou, W. *et al.* Exit from mitosis is triggered by Tem1-dependent release of the protein phosphatase Cdc14 from nucleolar RENT complex. *Cell* **97**, 233–244 (1999).
- Kamakaka, R. & Rine, J. Sir- and silencer-independent disruption of silencing in *Saccharomyces* by Sas10p. *Genetics* **149**, 903–914 (1998).
- Fournier, M. J. & Maxwell, E. S. The nucleolar snRNAs: catching up with the spliceosomal snRNAs. *Trends Biochem. Sci.* **18**, 131–135 (1993).
- Venema, J. & Tollervey, D. Processing of pre-ribosomal RNA in *Saccharomyces cerevisiae*. *Yeast* **11**, 1629–1650 (1995).
- Miller, O. L. Jr & Beatty, B. R. Visualization of nucleolar genes. *Science* **164**, 955–957 (1969).
- Mougey, E. B. *et al.* The terminal balls characteristic of eucaryotic rRNA transcription units in chromatin spread are rRNA processing complexes. *Genes Dev.* **7**, 1609–1619 (1993).
- Lee, S. J. & Baserga, S. J. Functional separation of pre-rRNA processing steps revealed by truncation of the U3 small nucleolar ribonucleoprotein component, Mpp10. *Proc. Natl Acad. Sci. USA* **94**, 13536–13541 (1997).
- Wehner, K. A. & Baserga, S. J. The σ^{70} -like motif: a eukaryotic RNA binding domain unique to a superfamily of proteins required for ribosome biogenesis. *Mol. Cell* **9**, 329–339 (2002).
- Warner, J. R. Nascent ribosomes. *Cell* **107**, 133–136 (2001).
- Knop, M. *et al.* Epitope tagging of yeast genes using a PCR-based strategy: more tags and improved practical routines. *Yeast* **15**, 963–972 (1999).
- Longtine, M. S. *et al.* Additional modules for versatile and economical PCR-based gene deletion and modification in *Saccharomyces cerevisiae*. *Yeast* **14**, 953–961 (1998).
- Schneider, B. L., Seufert, W., Steiner, B., Yang, Q. H. & Futcher, A. B. Use of polymerase chain reaction epitope tagging for protein tagging in *Saccharomyces cerevisiae*. *Yeast* **11**, 1265–1274 (1995).
- Ross-MacDonald, P. *et al.* Large-scale analysis of the yeast genome by transposon tagging and gene disruption. *Nature* **402**, 413–418 (1999).
- Dunbar, D. A., Dragon, F., Lee, S. J. & Baserga, S. J. A nucleolar protein related to ribosomal protein L7 is required for an early step in large ribosomal subunit biogenesis. *Proc. Natl Acad. Sci. USA* **97**, 13027–13032 (2000).
- Dunbar, D. A., Wormsley, S., Agentis, T. M. & Baserga, S. J. Mpp10p, a U3 small nucleolar ribonucleoprotein component required for pre-18S rRNA processing in yeast. *Mol. Cell. Biol.* **17**, 5803–5812 (1997).
- Lee, S. J. & Baserga, S. J. Imp3p and Imp4p: two specific components of the U3 small nucleolar ribonucleoprotein that are essential for pre-18S rRNA processing. *Mol. Cell. Biol.* **19**, 5441–5452 (1999).
- Hughes, J. M. X. & Ares, M. Jr Depletion of U3 small nucleolar RNA inhibits cleavage in the 5' external transcribed spacer of yeast pre-ribosomal RNA and prevents formation of 18S ribosomal RNA. *EMBO J.* **10**, 4231–4239 (1991).
- Samarsky, D. A. & Fournier, M. J. Functional mapping of the U3 small nucleolar RNA from the yeast *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* **18**, 3431–3444 (1998).
- Osheim, Y. N. & Beyer, A. L. Electron microscopy of RNP complexes on nascent RNA using the Miller chromatin spreading method. *Methods Enzymol.* **180**, 481–509 (1989).

Supplementary Information accompanies the paper on Nature's website (<http://www.nature.com/nature>).

Acknowledgements

We thank A. Djikeng, G. Dreyfuss, P. Gordon, J. Laney, T. Serio and T. Stone for assistance and advice. We also thank P. Glazer, M. Snyder and J. Steitz for critical reading of the manuscript. F.D. was supported by the Anna Fuller Fund for Molecular Oncology. This work was supported by federal grants from the NIH to D.F.H. and S.J.B. and the NSF to A.L.B., and by The Patterson Trust to S.J.B.

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to S.J.B. (e-mail: susan.baserga@yale.edu).

Structure of a bacterial quorum-sensing transcription factor complexed with pheromone and DNA

Rong-guang Zhang^{*}, Terina Pappas[†], Jennifer L. Brace^{*}, Paula C. Miller[‡], Tim Oulmassov[‡], John M. Molyneux[‡], John C. Anderson[‡], James K. Bashkin[‡], Stephen C. Winans[‡] & Andrzej Joachimiak^{*}

^{*} Bioscience Division/Structural Biology Center, Argonne National Laboratory, 9700 S. Cass Avenue, Argonne, Illinois 60439, USA

[†] Department of Microbiology, Cornell University, Ithaca, New York 14853, USA

[‡] Monsanto Company, 800 N. Lindbergh Boulevard Street, Louis, Missouri 63167, USA

[§] Present address: Pharmacia Corporation, St. Louis, Missouri, USA

Many proteobacteria are able to monitor their population densities through the release of pheromones known as *N*-acylhomoserine lactones. At high population densities, these pheromones elicit diverse responses that include bioluminescence, biofilm formation, production of antimicrobials, DNA exchange, pathogenesis and symbiosis¹. Many of these regulatory systems require a pheromone-dependent transcription factor similar to the LuxR protein of *Vibrio fischeri*. Here we present the structure of a LuxR-type protein. TraR of *Agrobacterium tumefaciens* was solved at 1.66 Å as a complex with the pheromone *N*-3-oxooctanoyl-L-homoserine lactone (OOHL) and its TraR DNA-binding site. The amino-terminal domain of TraR is an $\alpha/\beta/\alpha$ sandwich that binds OOHL, whereas the carboxy-terminal domain contains a helix–turn–helix DNA-binding motif. The TraR dimer displays a two-fold symmetry axis in each domain; however, these two axes of symmetry are at an approximately 90° angle, resulting in a pronounced overall asymmetry of the complex. The pheromone lies fully embedded within the protein with virtually no

solvent contact, and makes numerous hydrophobic contacts with the protein as well as four hydrogen bonds: three direct and one water-mediated.

Many bacteria communicate by releasing specific chemical signals (pheromones) whose concentrations increase at high population densities. Bacteria are thought to use these signals to induce expression of particular target genes only at high population sizes, a phenomenon referred to as quorum sensing¹. A large family of regulatory systems has been described, each resembling the LuxI and LuxR proteins of *V. fischeri*. LuxI-type proteins synthesize *N*-acylhomoserine lactone pheromones (AHLs, also referred to as autoinducers), which diffuse from the bacteria that produce them either passively or by means of active efflux, and accumulate at high population densities. AHLs are thought to bind to and activate LuxR-type receptor proteins, which function as cytoplasmic transcription factors, either as transcriptional activators or repressors².

LuxR is composed of two domains. An *Escherichia coli* strain that overexpresses the N-terminal fragment of LuxR sequesters its cognate AHL, indicating that this domain is sufficient for pheromone binding³. Overexpression of this fragment also has a dominant negative effect on the activity of the full-length protein, suggesting that this domain can mediate protein oligomerization⁴. The C-terminal domain of LuxR-type proteins has a predicted helix–turn–helix (HTH) motif and is thought to make sequence-specific contacts with DNA. The C-terminal domain, when overexpressed *in vivo*, is sufficient to activate transcription⁵, and has been shown *in vitro* to facilitate RNA polymerase binding to a target promoter⁶. Studies of LuxR-type proteins have been hindered by a lack of structural data.

TraR of *Agrobacterium tumefaciens* is a close homologue of LuxR and displays similar properties. The protein is coded for by the tumour-inducing (Ti) plasmid of this plant-pathogenic bacterium, where it activates *tra* genes, required for interbacterial plasmid transfer⁷. The *traR* gene is induced by nutrients called opines that are released from plant tumours, ensuring that quorum sensing by *A. tumefaciens* occurs only in the tumour environment. Both LuxR

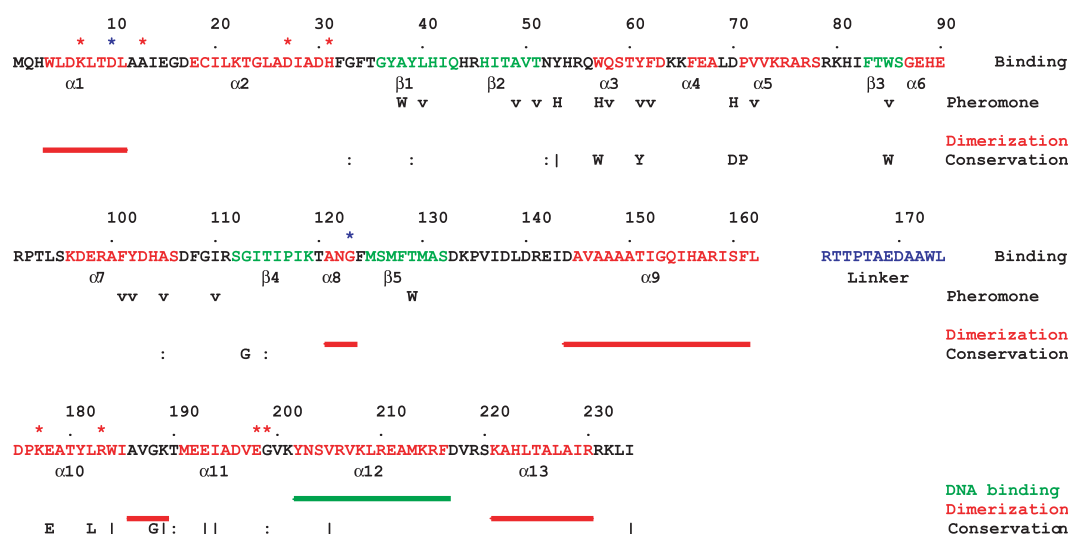


Figure 1 Functional and structural roles of amino acid residues in TraR. Residue colour scheme: red, helical; green, form β -strands; blue, indicate an inter-domain linker; black, random coils. Residues participating in pheromone binding are indicated using the letters v (van der Waals interactions), H (direct protein–pheromone hydrogen bonding), and W (water-mediated hydrogen bonding). Residues that participate in protein dimerization are marked with a red bar, whereas the DNA binding helix is indicated with a green bar.

Residues that are identical in at least 95% of LuxR-type proteins are indicated using a letter, whereas residues that are at least 80% or 60% identical in LuxR-type proteins are marked with a vertical line or a colon, respectively (see ref. 1 for a recent compilation). Residues required for positive control are marked with a blue asterisk; those involved in N- and C-terminal domain interactions are marked with a red asterisk.

and TraR require their respective pheromones for DNA binding *in vivo*^{8,9}. The TraR N-terminal domain acts as a transdominant mutant and inactivates the full-length protein *in vivo* and *in vitro* by forming heterodimers^{9,10}. We previously purified TraR as a complex with the pheromone OOHL bound at a one-to-one stoichiometry¹¹. TraR is a dimer in solution and binds target promoters *in vitro* as a dimer^{12,13}. TraR recognizes specifically 18-base pair (bp) palindromic DNA sequences called 'tra boxes', one of which is centred 42 bp upstream of the divergent *traA* and *traC* promoters¹¹. A similar *tra* box is found upstream of the *traI* promoter, which is regulated by TraR⁷.

We crystallized TraR in the presence of OOHL and a self-complementary oligonucleotide containing the canonical *tra* box sequence. The crystal asymmetric unit contains two structurally similar TraR dimers, each binding two molecules of OOHL and one duplex DNA fragment. The N-terminal domain of each monomer (residues 1–162) binds one molecule of OOHL, whereas the C-terminal domain (residues 175–234) binds to half of a *tra* box. A twelve-residue amino acid linker joins these domains (Fig. 1, residues indicated in blue). The two N-terminal domains and OOHL molecules of the TraR dimer show a two-fold rotational symmetry, and similarly, the two C-terminal domains and bound *tra* box DNA also show a two-fold rotational symmetry. However, the two axes of rotation are at an approximately 90° angle to each other (red lines in Fig. 2), creating a pronounced asymmetry for the entire complex (see below).

The N-terminal domain consists of an $\alpha/\beta/\alpha$ sandwich. The central β -sheet has five anti-parallel strands and is curved, with the pheromone bound to its concave surface. The pheromone is fully embedded within the protein and has no significant contact with solvent (Figs 2 and 3a). Hydrophobic and aromatic residues constitute most of the OOHL binding pocket. The fatty acyl chain runs parallel to the β -sheet surface and is accommodated by several hydrophobic side chains (Leu 40, Thr 51, Tyr 53, Tyr 61, Phe 62, Val 72, Trp 85, Ile 110). Several of these residues are highly conserved among LuxR family members (Tyr 53, Tyr 61, Val 72, Trp 85, Ile 110) and make extensive van der Waals contacts with the pheromone. In addition, the highly conserved residue Gly 113 positions the side chain of Trp 85. Hydrogen bonds contribute to pheromone binding, including one between Trp 57 and the ring keto group, another between Asp 70 and the imino group, and another between Tyr 53 and the 1-keto group (Fig. 3b). A fourth hydrogen bond is made between the 3-keto group of OOHL and a water molecule that is

stabilized by hydrogen bonding to Thr 129 and Ala 38 (Fig. 3b). All of these residues are highly conserved¹.

The complete encapsulation of the pheromone by TraR is similar to the binding of the insect pheromone bombykol to its cognate receptor protein¹⁴. This encapsulation helps to explain why OOHL is bound virtually irreversibly to TraR^{11,13}. We have previously shown that TraR synthesized in *E. coli* in the absence of OOHL is proteolysed rapidly by the Clp and Lon proteases, and that its half-life both in *E. coli* and in *A. tumefaciens* is extended 20-fold by the addition of OOHL^{11,12}. These experiments indicate that TraR may require OOHL as a scaffold to help it acquire a protease-resistant tertiary structure. However, this is probably not true for all regulatory system members, as some LuxR-type proteins form multimers even in the absence of pheromone¹⁵ and EsaR protein is thought to bind DNA only in the absence of pheromone¹⁶. Furthermore, these DNA-bound apoproteins remain receptive to pheromone, indicating that their pheromone binding sites must be accessible.

The DNA-binding domain of TraR is a four-helix bundle containing a HTH DNA-binding motif⁷. The DNA in the complex has a canonical B conformation and is slightly bent, following the surface of the protein (Fig. 2). We confirmed this DNA bend by gel mobility assays using permuted DNA fragments containing *tra* box DNA (data not shown). TraR recognizes the *tra* box DNA primarily through a recognition helix (residues 202–216) that lies perpendicular to the DNA axis and penetrates the surface of the major groove (Fig. 4). The orientation of the recognition α -helix is stabilized by a salt bridge formed between the conserved residues Arg 215 and Glu 178. TraR makes sequence-specific contacts with six bases in each half-site of the *tra* box. Sequence-specific hydrogen bonds are made primarily by Arg 206 and Arg 210. Arg 206 contacts G₋₅ and C₋₄ of one DNA strand and G₄ of the complementary strand. Arg 206 also forms a hydrogen bond with Asn 203, which in turn contacts phosphate₋₇, while Arg 210 contacts G₋₇, T₋₆ and G₋₅ of one strand and C₅ of the complementary strand. Additional van der Waals contacts are made between Val 207 and G₋₇, Val 207 and T₋₈, and Glu 211 and T₋₈. Tyr 202 makes a water-mediated hydrogen bond with T₃ and hydrogen bonds with Arg 206. Residues Met 191 (main chain imino group), Asn 203, Ser 204, Lys 208 and Lys 221 contact the DNA backbone. The completely conserved residue Gly 188 is a part of the DNA-binding domain dimerization interface. It permits the close approach of the protein to the minor groove, although no contact is made with DNA bases.

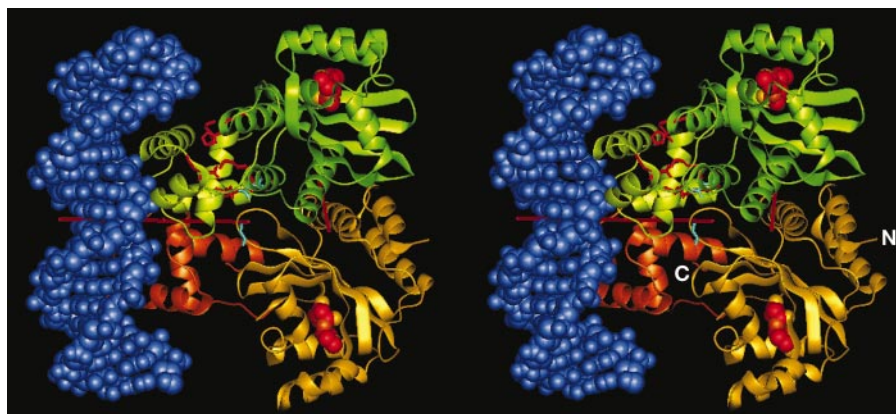
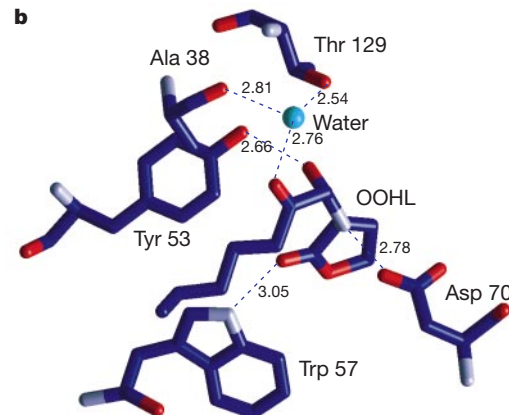


Figure 2 Stereo view of the structure of the TraR–OOHL–DNA complex. Domains in the two monomers are shown in different colours (light/dark orange and light/dark green), whereas the DNA is coloured blue and the OOHL is coloured red. Note that the two-fold dyad axis of the DNA and DNA-binding domains lies in the plane of the page (horizontal red line), whereas that relating to the pheromone-binding domains is swiveled by

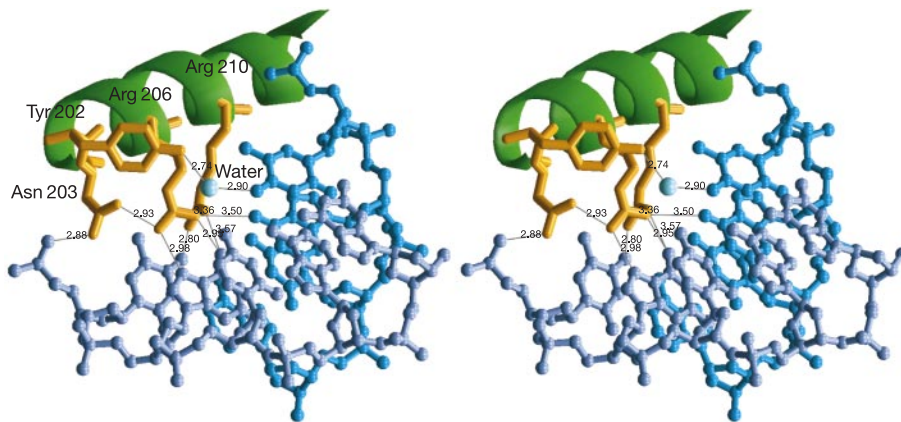
approximately 90° (short red line). Side chains of residues in the upper monomer (light/dark green) that mediate interaction between DNA-binding and pheromone-binding domains are shown in red and residues that affect transcription activation are shown in light blue. The N terminus and C terminus of the lower subunit are labelled.



The hydrogen bond between the 3-keto group and protein is water-mediated. The distance between interacting atoms is shown in Å.

the N- and C-terminal domains is disordered. In contrast, the second monomer shows almost no contacts between the N- and C-terminal domains.

The TraR protein of the nopaline-type Ti plasmid pTiC58 (over 90% identical to TraR of the octopine-type Ti plasmid) has been subjected to extensive mutagenesis. Deletion of the N-terminal 2–4 residues abolishes transcriptional activity and has dominant negative effects *in vivo*⁹. Our structure shows that residue Trp 4 is buried within the domain and may contribute to structural integrity. Deletion of two amino acids from the C terminus abolishes *trab*-box binding *in vivo* and causes a strong dominant negativity against wild-type TraR⁹. Both side chains are buried within the domain, forming part of the hydrophobic core, and may contribute to the integrity of this domain. Similarly, the double mutations M213I-R215H and M191I-R206H also block DNA binding *in vivo*⁹. Arg 215 forms a salt bridge with Glu 178, whereas, as described above, Arg 206 makes sequence-specific DNA contacts. The C-terminal domain of TraR also binds with high affinity to the antiactivator protein TraM¹⁹. Residues Pro 176, Leu 182 and Ala 195 were reported to be important for this interaction. Structural analysis of this region shows that these residues are scattered throughout the domain and are unlikely to contact TraM simultaneously¹⁹. Residues Asp 10 and Gly 123 are essential for transcription activation but not for specific DNA binding⁹, suggesting that these residues could make specific contact with RNA polymerase (RNAP). Both



molecule (yellow) forms a hydrogen bond between the DNA and Tyr 202. Distances are shown in Å.

residues are exposed to solvent, and Asp 10 of each subunit is closely juxtaposed to Gly 123 of the opposite subunit. Owing to the asymmetry of the complex, one of the Asp 10–Gly 123 pairs lies close to the DNA-binding domain (and could contact RNAP), whereas the other Asp 10–Gly 123 pair lies far away from the DNA-binding domain. Mutational studies of LuxR suggest that residues in the DNA-binding domain are involved in interactions with RNAP²⁰.

Unexpectedly, TraR is strongly asymmetric. In the asymmetric TraR dimer the DNA-binding C-terminal domains and pheromone-binding N-terminal domains behave as distinct units, and the asymmetry of the complex is instigated by a small number of contacts between pheromone- and DNA-binding domains. These interactions are sufficient to maintain a specific orientation in the crystal, as the crystal lattice contacts are predominantly between the N-terminal domains and DNA. The asymmetry of the TraR dimer is reminiscent of that of NarL, whose N-terminal receiver domain contacts one side of the C-terminal DNA-binding domain¹⁸. The asymmetry of NarL was presented as confirmatory evidence that the N-terminal domain inhibits the function of the C-terminal domain. In contrast, as the TraR crystals described here contain OOH and *tra* box DNA, TraR must be in an active conformation. TraR could show the same asymmetry *in vivo*, and it also seems likely that the observed asymmetry reflects the conformational freedom between domains allowing TraR to function on bi-directional promoters *in vivo*.

We describe here the structure of a functional complex of quorum-sensing transcriptional regulator TraR from *A. tumefaciens*, a member of LuxR family. Our structural studies support earlier work that TraR requires OOH as a scaffold for folding and dimerization^{11,12}. The pheromone-binding domains appear to provide an extensive dimerization interface, while the DNA-binding domains contribute additional interactions. As a result the pheromone indirectly affects gene activation by increasing stability of TraR and the formation of functional dimers that are predisposed to decode specific TraR-binding sites and activate transcription. The TraR–OOH–*tra* DNA ternary complexes described here should serve as a prototype for the large family of AHL-induced transcription activators. □

Methods

TraR production and purification

TraR was purified from *A. tumefaciens* strain B21(DE3) harbouring pJZ358 (ref. 11) cultured in LB broth supplemented with 5 μ M OOH. TraR production was induced using 200 μ M isopropyl- β -D-thiogalactopyranoside. Cells were suspended in a buffer containing 10 mM Na₂HPO₄ (pH 7.0), 150 mM NaCl, 0.1 mM EDTA, 1 mM dithiothreitol 5% glycerol, and 100 mM OOH, and lysed using a French Press. The resulting cleared lysate was purified by successive heparin Sepharose, SP-Sepharose, and MonoS chromatography columns (Pharmacia). TraR was eluted from each using linear gradients from 150 mM to 1 M NaCl. Selenomethionyl (SeMet) TraR was purified as described above from strain B834(DE3) harbouring pJZ358. This strain is a derivative of BL21(DE3), which is auxotrophic for methionine. The strain was cultured in AT minimal medium²¹ supplemented with 10 mg l⁻¹ thiamine, 40 mg l⁻¹ L-selenomethionine (Sigma), and 5 μ M OOH.

Crystallization, structure determination and refinement

The DNA constructs for co-crystallization were designed using rules described previously²². Crystals were produced by vapour diffusion in the presence of the self-complementary oligonucleotide GATGTGCAGATCTGCACATC (underlined residues represent the canonical 18-nucleotide TraR-binding site; nucleotides run from –10 through to 10 (left to right, as in text)). These crystals belong to space group P2₁ and diffract X-rays to 1.66 Å resolution using synchrotron radiation. A three-wavelength multiple anomalous dispersion (MAD) data set was collected to 1.9 Å resolution at the Structural Biology Center 19ID beamline using a protocol described earlier^{23,24}. Data were processed and scaled with the HKL2000 software package²⁵. A heavy atom search using CNS²⁶ yielded 8 of the 12 Se sites predicted from the sequence, and these sites were used in phasing; the figure of merit was 0.468. The electron density was calculated to 1.9 Å resolution. After density modification, the DNA region, protein region, pheromone, and ordered solvent were fully interpretable. Protein model building was carried out using a semi-automated approach. The initial model (90%) was built with wARP²⁷. Missing sections and poorly fitted regions of the protein were built or corrected manually using QUANTA²⁸. The DNA structure and the pheromone model were also built manually using

QUANTA. After most of the model was built, unassigned electron density was found in the interior of the N-terminal domain and was assigned to bound pheromone. The structure was refined with CNS with a crystallographic *R*-factor of 23.0% and *R*_{free} value of 25.2%.

Received 23 January; accepted 18 April 2002; doi:10.1038/nature00833.

- Whitehead, N. A., Barnard, A. M. L., Slater, H., Simpson, N. J. L. & Salmond, G. P. C. Quorum-sensing in Gram-negative bacteria. *FEMS Microbiol. Rev.* **25**, 365–404 (2001).
- Fuqua, C., Parsek, M. R. & Greenberg, E. P. Regulation of gene expression by cell-to-cell communication: acyl-homoserine lactone quorum sensing. *Annu. Rev. Genet.* **35**, 439–468 (2001).
- Hanzelka, B. L. & Greenberg, E. P. Evidence that the N-terminal region of the *Vibrio fischeri* LuxR protein constitutes an autoinducer-binding domain. *J. Bacteriol.* **177**, 815–817 (1995).
- Choi, S. H. & Greenberg, E. P. Genetic evidence for multimerization of LuxR, the transcriptional regulator of *Vibrio fischeri* luminescence. *Mol. Mar. Biol. Biotechnol.* **1**, 408–413 (1992).
- Choi, S. H. & Greenberg, E. P. The C-terminal region of the *Vibrio fischeri* LuxR protein contains an inducer-independent lux gene-activating domain. *Proc. Natl Acad. Sci. USA* **88**, 1115–1119 (1991).
- Stevens, A. M., Dolan, K. M. & Greenberg, E. P. Synergistic binding of the *Vibrio fischeri* LuxR transcriptional activator domain and RNA polymerase to the lux promoter region. *Proc. Natl Acad. Sci. USA* **91**, 12619–12623 (1994).
- Fuqua, W. C. & Winans, S. C. A LuxR–LuxI type regulatory system activates *Agrobacterium* Ti plasmid conjugal transfer in the presence of a plant tumour metabolite. *J. Bacteriol.* **176**, 2796–2806 (1994).
- Egland, K. A. & Greenberg, E. P. Conversion of the *Vibrio fischeri* transcriptional activator, LuxR, to a repressor. *J. Bacteriol.* **182**, 805–811 (2000).
- Luo, Z. Q. & Farrand, S. K. Signal-dependent DNA binding and functional domains of the quorum-sensing activator TraR as identified by repressor activity. *Proc. Natl Acad. Sci. USA* **96**, 9009–9014 (1999).
- Chai, Y., Zhu, J. & Winans, S. C. A defective TraR-like protein of *Agrobacterium tumefaciens* forms heterodimers with TraR *in vitro*, thereby blocking TraR-mediated quorum sensing. *Mol. Microbiol.* **40**, 414–421 (2001).
- Zhu, J. & Winans, S. C. Autoinducer binding by the quorum-sensing regulator TraR increases affinity for target promoters *in vitro* and decreases traR turnover rates in whole cells. *Proc. Natl Acad. Sci. USA* **96**, 4832–4837 (1999).
- Luo, Z. Q. & Winans, S. C. The quorum-sensing regulator TraR of *Agrobacterium tumefaciens* requires autoinducer for protein folding, dimerization, and protease resistance. *Proc. Natl Acad. Sci. USA* **98**, 1507–1512 (2001).
- Qin, Y. *et al.* Quorum-sensing signal binding results in dimerization of TraR and its release from membranes into the cytoplasm. *EMBO J.* **19**, 5212–5221 (2000).
- Sandler, B. H., Nikonova, L., Leal, W. S. & Clardy, J. Sexual attraction in the silkworm moth: structure of the pheromone-binding-protein-bombykol complex. *Chem. Biol.* **7**, 143–151 (2000).
- Welch, M. *et al.* N-acyl homoserine lactone binding to the CarR receptor determines quorum-sensing specificity in *Erwinia*. *EMBO J.* **19**, 631–641 (2000).
- von Bodman, S. B., Majerczak, D. R. & Coplin, D. L. A negative regulator mediates quorum-sensing control of exopolysaccharide production in *Pantoea stewartii* subsp. *Proc. Natl Acad. Sci. USA* **95**, 7687–7692 (1998).
- Nelson, H. C. Structure and function of DNA-binding proteins. *Curr. Opin. Genet. Dev.* **5**, 180–189 (1995).
- Baikalov, I. *et al.* NarL dimerization? Suggestive evidence from a new crystal form. *Biochemistry*, **37**, 3665–3676 (1998).
- Luo, Z.-Q., Qin, Y. & Farrand, S. K. The antiactivator TraM interferes with the autoinducer-dependent binding of TraR to DNA by interacting with the C-terminal region of the quorum-sensing activator. *J. Biol. Chem.* **275**, 7713–7722 (2000).
- Egland, K. A. & Greenberg, E. P. Quorum sensing in *Vibrio fischeri*: analysis of the LuxR DNA binding region by alanine-scanning mutagenesis. *J. Bacteriol.* **183**, 382–386 (2001).
- Tempe, J., Petit, A., Holsters, M., Van Montagu, M. & Schell, J. Thermosensitive step associated with transfer of the Ti plasmid during conjugation: possible relation to transformation in crown gall. *Proc. Natl Acad. Sci. USA* **74**, 2848–2849 (1977).
- Joachimski, A. & Sigler, P. B. Crystallization of protein–DNA complexes. *Methods Enzymol.* **208**, 82–99 (1991).
- Walsh, M. A., Dementieva, I., Evans, G., Sanishvili, R. & Joachimski, A. Taking MAD to the extreme: ultrafast protein structure determination. *Acta Crystallogr. D* **55**, 1168–1173 (1999).
- Walsh, M. A., Evans, G., Sanishvili, R., Dementieva, I. & Joachimski, A. MAD data collection—current trends. *Acta Crystallogr. D* **55**, 1726–1732 (1999).
- Otwinski, Z. & Minor, W. Processing of x-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307–326 (1997).
- Brünger, A. *et al.* Crystallography & NMR system: a new software suite for macromolecular structure determination. *Acta Crystallogr. D* **54**, 905–921 (1998).
- Perrakis, A., Morris, R. & Lamzin, V. S. Automated protein model building combined with iterative structure refinement. *Nature Struct. Biol.* **6**, 458–463 (1999).
- QUANTA, Molecular Simulations Inc, San Diego. (2000).

Acknowledgements

This work was supported by Monsanto Company, the US Department of Energy, Office of Biological and Environmental Research, and a National Research Service Award to S.C.W.

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to A.J.

(e-mail: andrzej@anl.gov). The atomic coordinates have been deposited with Research Collaboratory for Structural Bioinformatics (RCSB) into the Protein Data Bank (PDB) under code 1L3L for the PDB and 015611 for RCSB.

Structural basis for the recognition of hydroxyproline in HIF-1 α by pVHL

Wai-Ching Hon[†], Michael I. Wilson^{†‡§}, Karl Harlos^{*}, Timothy D. W. Claridge^{||}, Christopher J. Schofield^{||}, Christopher W. Pugh[‡], Patrick H. Maxwell[‡], Peter J. Ratcliffe[‡], David I. Stuart^{*} & E. Yvonne Jones^{*}

^{*} Division of Structural Biology, and [‡] Cellular Physiology Group, Henry Wellcome Building of Genomic Medicine, Roosevelt Drive, Oxford OX3 7BN, UK ^{||} Oxford Centre for Molecular Sciences and Dyson Perrins Laboratory, South Parks Road, Oxford OX1 3QY, UK

[†] These authors contributed equally to this work

[§] Present address: Division of PNAC, MRC Laboratory of Molecular Biology, Hills Road, Cambridge CB2 2QH, UK

Hypoxia-inducible factor-1 (HIF-1) is a transcriptional complex that controls cellular and systemic homeostatic responses to oxygen availability¹. HIF-1 α is the oxygen-regulated subunit of HIF-1, an $\alpha\beta$ heterodimeric complex¹. HIF-1 α is stable in hypoxia, but in the presence of oxygen it is targeted for proteasomal degradation by the ubiquitination complex pVHL, the protein of the von Hippel–Lindau (VHL) tumour suppressor gene and a component of an E3 ubiquitin ligase complex^{2,3}. Capture of HIF-1 α by pVHL is regulated by hydroxylation of specific prolyl residues in two functionally independent regions of HIF-1 α ^{4–7}. The crystal structure of a hydroxylated HIF-1 α peptide bound to VCB (pVHL, elongins C and B) and solution binding assays reveal a single, conserved hydroxyproline-binding pocket in pVHL. Optimized hydrogen bonding to the buried hydroxyprolyl group confers precise discrimination between hydroxylated and unmodified prolyl residues. This mechanism provides a new focus for development of therapeutic agents to modulate cellular responses to hypoxia.

Ubiquitin-mediated degradation of specific proteins is central in the regulation of many cellular events⁸. Such processes require mechanisms for specific recognition of proteolytic targets by E3 ubiquitin ligases. Several multisubunit, cullin-containing, RING E3 ubiquitin ligases use phosphorylation to regulate capture of protein substrates⁹. In contrast, recognition of HIF-1 α by VCB E3 ligase (pVHL, elongins B and C, Cul2 and Rbx1; ref. 10) requires hydroxylation of specific prolyl residues by a group of 2-oxogluta-

rate-dependent, non-haem dioxygenases^{11,12}. In human HIF-1 α , hydroxylation of two functionally independent prolyl residues, 402 and 564—which are within amino- and carboxy-terminal oxygen-dependent degradation (NODD and CODD) motifs^{7,13,14}—causes direct and highly efficient interaction with pVHL. We here address how the insertion of a single oxygen atom into a prolyl residue governs selective recognition, by examining the crystal structure of human VCB in complex with a peptide containing the human HIF-1 α CODD region (residues 549–582) that has a 4(R)-L-hydroxyproline (Hyp) as residue 564.

The structure was determined by molecular replacement using the apo form of VCB¹⁵ as a search model. Reflection data from 30 to 2.0 Å were used for model rebuilding and refinement. Electron density for HIF-1 α peptide residues 560–577 is of good quality, with only the side chain of E560 poorly defined (Fig. 1). The peptide lies in an extended conformation across one side of the β -domain of pVHL (Fig. 2a, b), binding to one of the β -sheets as if it were a complementary β -strand, although there is no extensive main-chain hydrogen bonding to define it as such. The binding does not involve conformational changes to the main chain of pVHL (r.m.s. deviation of C α atoms is 0.3 Å between the apo and peptide-bound forms of the β -domain). There is some rearrangement of the three VCB subunits compared to the apo structure (reflecting differing crystal contacts), but the alterations are distant from the CODD binding site.

Within the CODD peptide, residues 560–567 and 571–577 form distinct binding sites (sites 1 and 2, respectively). Between these sites only the side chain of D569 contacts pVHL. The binding surface on pVHL spans residues 67–117 of the β domain (Fig. 2a–c). The buried surface area (1,850 Å²) and the shape complementarity¹⁶ (0.79) between CODD and pVHL are comparable to those of the inter-subunit interfaces in VCB (2,050 and 2,300 Å²; 0.74 and 0.71). The pVHL–peptide interactions are predominantly nonpolar but also include ten electrostatic bonds (Figs 2c, d, and 3a, b).

Site 1 (EMLAHypYIP) is the primary binding site, contributing more than 60% of the total buried surface area and van der Waals contacts. It contains the leucine–alanine–proline (LAP) motif found in HIF α -subunits, where P is converted to Hyp^{4–7,12} (Fig. 2d). Two residues, HypY, contribute all five hydrogen bonds in this site. The peptide-binding cavity on pVHL is, in general, rather shallow, with LAHypYIP following the curvature of the binding surface (Fig. 3a). The binding surface for site 2 (DFQLRSF) is relatively flat and only one residue of the peptide makes specific side-chain interactions (Fig. 3b), arguing against extensive sequence

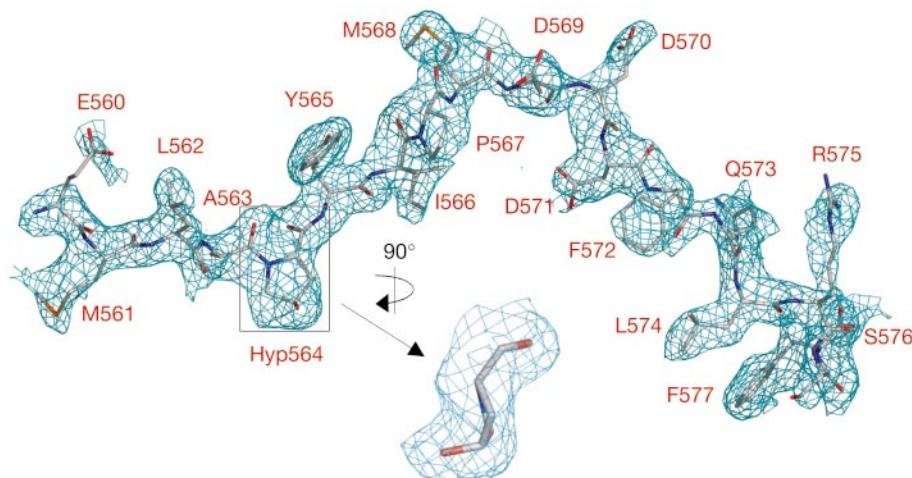


Figure 1 Electron density for the bound CODD peptide. The boomerang-shaped CODD peptide is shown as a stick model (where the carbon, nitrogen, oxygen and sulphur atoms are grey, blue, red and yellow, respectively). The difference electron density (contoured at

1.5 σ) was calculated using phases derived from the model after refinement by simulated annealing, with CODD omitted to remove model bias.

specificity in this region. R575 does not bind pVHL, splitting site 2 into two subsites. Five electrostatic bonds are contributed by the subsite DFQL; two are salt bridges formed by D571, and the rest are hydrogen bonds involving the main-chain atoms of F572 and L574. FQL is conserved as a nonpolar–polar–nonpolar motif: LDF in the NODD motif and MQM in the *Caenorhabditis elegans* ODD motif (Fig. 2d).

Hyp564 is almost entirely buried in a deep pocket in the site 1 binding cavity (Fig. 3c), contributing 20% of the total van der Waals contacts between the peptide and pVHL. The shape of this pocket (the only side-chain-specific cavity in the CODD-binding region of pVHL) complements precisely the up-pucker conformation of the hydroxyproline ring. The pocket is lined by W88, Y98, S111, H115 and W117, which are invariant between human and *C. elegans* pVHL (except for S111, which is conservatively substituted by threonine) (Fig. 2c). Residues S111 and H115 are sunk within the β -domain, with the exposed S111 hydroxyl group and H115 imidazole amino group (atom N3/N δ) forming the floor of the Hyp-binding pocket and serving as hydrogen-bonding partners to the Hyp564 hydroxyl group (Fig. 3a, c). These two hydrogen bonds are of optimal distances and geometry (Fig. 3d). The positions of the five Hyp-binding residues are conserved between the apo and peptide-bound structures, locked by interdependent side-chain packing. In addition, the side-chain orientation of H115 is fixed by van der Waals contacts with S65, Y112, R113 and G114. H115 is part of a water-mediated hydrogen-bonding network (Fig. 3d), involving a water molecule buried in a pocket neighbouring the Hyp-binding pocket (Fig. 3a, c). This water is also present in the apo structure. It coordinates the H115 imidazole amino group distal to Hyp564, the side chain of N67 (which is functionally conserved as a histidine in *C. elegans* pVHL; Fig. 2c), the main-chain amide nitrogen of R69 and the carbonyl oxygen of L562 (CODD). By

hydrogen bonding to the opposed imidazole amino group of H115, the water molecule stabilizes the Hyp-binding imidazole amino group as a conjugate base. The resulting partial charge may strengthen the hydrogen bond, rendering H115 the hydrogen bond acceptor for Hyp564, with S111 acting as a hydrogen bond donor. The S111 hydroxyl group is in turn a hydrogen bond acceptor of the indole nitrogen of W117. Hence Hyp564 completes an elaborate network of hydrogen bonds, providing a chemical basis for efficient capture of HIF-1 α after oxygen-dependent hydroxylation. Although the physiological importance of this process is established, its relationship to pVHL tumour suppressor function is less clear. It is noteworthy that all five pVHL residues lining the Hyp-binding pocket are affected by missense mutations in VHL disease¹⁷ (Fig. 2c), suggesting that failure to capture HIF-1 α , and/or another hydroxylated target, is important to the tumour-promoting mechanism associated with VHL disease.

Surface plasmon resonance data were measured for VCB binding to a range of peptides (Table 1 and Supplementary Information). Hydroxylated human CODD and NODD, and *C. elegans* ODD peptides, bind VCB with similar kinetics and binding affinities (29–53 nM). Moreover, competition studies indicate that human CODD and NODD bind to the same site. Both hydroxylated human CODD and a short LAHypYIP peptide derived from site 1 inhibit binding of VCB to human NODD (half-maximal inhibitory concentration, IC₅₀, about 1.0 and 1.3 μ M, respectively). Peptides of human CODD and NODD containing just site 1 have similar binding kinetics and only slightly weaker binding than their counterparts containing sites 1 and 2, demonstrating that site 1 drives the binding, as expected from the crystal structure.

The similar affinity of pVHL for the very different ODD peptides is consistent with only Hyp occupying a side-chain-specific pocket. Binding of both CODD and NODD to a single pVHL site is also

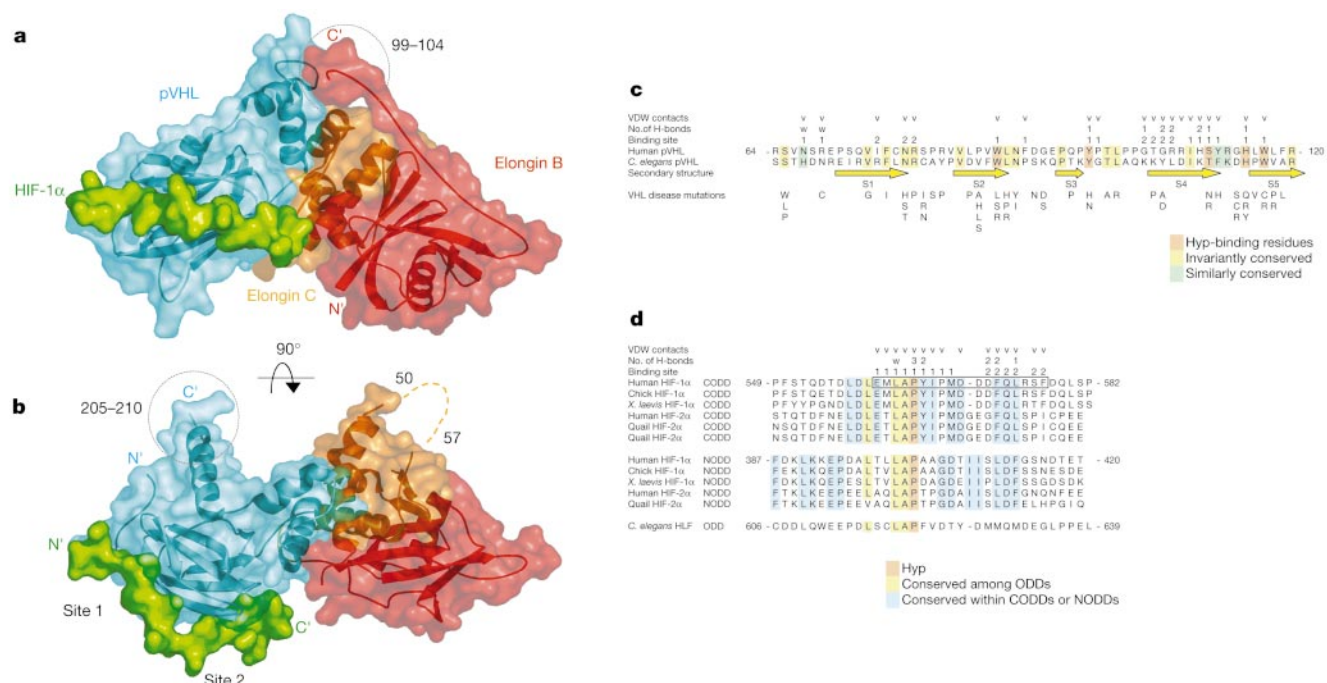


Figure 2 Structure of VCB–CODD complex, and sequence alignments of HIF ODD regions and pVHL. **a, b**, Surface rendering of the VCB–CODD complex with the secondary structures revealed underneath the semitransparent surface. Two regions seen in the current structure but not included in the previously reported model of apo VCB¹⁵ are circled in elongin B (residues 99–104) and pVHL (residues 205–210). The C-terminal tail of elongin B makes newly observed contacts with the α domain (V170 and N174) of pVHL. The still-unobserved region (residues 50–57) in elongin C is modelled as a dotted loop.

c, d, Sequence alignments of the CODD-binding region of pVHL β domain (**c**) and of the ODD motifs of HIF homologues (**d**). Residues that form van der Waals (VDW) contacts are indicated by 'v'. 'w' denotes single hydrogen bonds (H-bonds) formed with a key water molecule (see Fig. 3d), as opposed to those formed between pVHL and CODD (see Fig. 3a, b). Arrows denote β -strands. VHL disease missense mutations¹⁷ that occur in the CODD-binding region of pVHL are listed.

Table 1 Binding kinetics of VCB to HIF-1 α peptides as assayed by surface plasmon resonance

Ligand	Sequence	k_a ($\times 10^5 \text{ M}^{-1} \text{ s}^{-1}$)	k_d ($\times 10^{-2} \text{ s}^{-1}$)	K_D ($\times 10^{-8} \text{ M}$)
Sites 1 + 2				
BT-hCDD34OH	549-PFSTQDTLDLEMLAPYIPMD-DDFQLRSFDQLSP	7.0 ± 0.02	2.2 ± 0.01	3.1
BT-hCDD19OH	556-DLDLEMLAPYIPMD-DDFQL	8.4 ± 0.06	2.7 ± 0.03	3.3
BT-hNODD28OH	389-LKKEPDALTLLAPAGDTIISLDFGSND	7.9 ± 0.07	4.1 ± 0.03	5.2
BT-ceODD28OH	607-DDLQWEEPDLSCLAFFVDITY-DMMQMDEG	7.6 ± 0.03	2.2 ± 0.01	2.9
Site 1				
BT-PEO-Cys-hCDD12OH	558-DLEMLAPYIPMD	6.6 ± 0.05	4.2 ± 0.02	6.4
BT-PEO-Cys-hNODD12OH	393-DALTLLAPAGD	4.9 ± 0.02	4.5 ± 0.01	9.3

BT, biotin; h, human; ce, *C. elegans*; PEO, biotin-PEO-maleimide linker; OH denotes presence of Hyp; underlined P is Hyp. Only the first residues of the peptides are numbered. k_a (on-rate) and k_d (off-rate) are the kinetic association and dissociation constants, respectively, for a 1:1 Langmuir binding model. The equilibrium dissociation constant (affinity constant), $K_D = k_d/k_a$. Regions corresponding to sites 1 and 2 are shown in bold.

consistent with *in vivo* observations that each pVHL binding motif on HIF-1 α can function in isolation, and that overexpression of either peptide can overcome endogenous HIF-1 α degradation (C. Willam and C.W.P., unpublished results).

No stable complexes could be detected between VCB and unhydroxylated CODD for protein concentrations up to 45 μM . The steady-state affinity constant was estimated to be 34 μM (see Supplementary Information), which is about 1,000-fold less than the affinity for hydroxylated CODD. In principle, this discrimination could derive from specific binding effects and/or from conformational changes in either the pVHL or peptides. The crystallographic results show no significant conformational changes in pVHL on binding the high-affinity peptide, and ^1H -nuclear magnetic resonance (NMR) experiments on the hydroxylated and unmodified peptides indicate that they adopt similar conformations in solution (see Supplementary Information). Apart from minor entropic effects due to differences in the torsional

properties of proline and hydroxyproline, all of the discrimination must arise, therefore, from the interaction with pVHL. The drastic decrease in binding affinity owing to the absence of a hydroxyl group translates to about 4 kcal mol^{-1} of free energy change, which is consistent with the loss of two hydrogen bonds. Indeed, Hyp binding is stabilized by two hydrogen bonds, which would be denied to proline, and although proline could fit into the Hyp-binding pocket, it would exclude a water molecule that hydrogen bonds to the rigidly positioned S111 and H115 in the ligand-lacking structure. There would therefore be a net loss of two hydrogen bonds on binding the unhydroxylated peptide compared with the hydroxylated peptide. Hence, the specificity for Hyp is based on the special positions and environment of its hydrogen-bonding partners. This mechanism is reminiscent of that used by tyrosyl transfer RNA synthetase to achieve similar discrimination between tyrosine and phenylalanine¹⁸.

The central role of the HIF/pVHL system in a range of responses

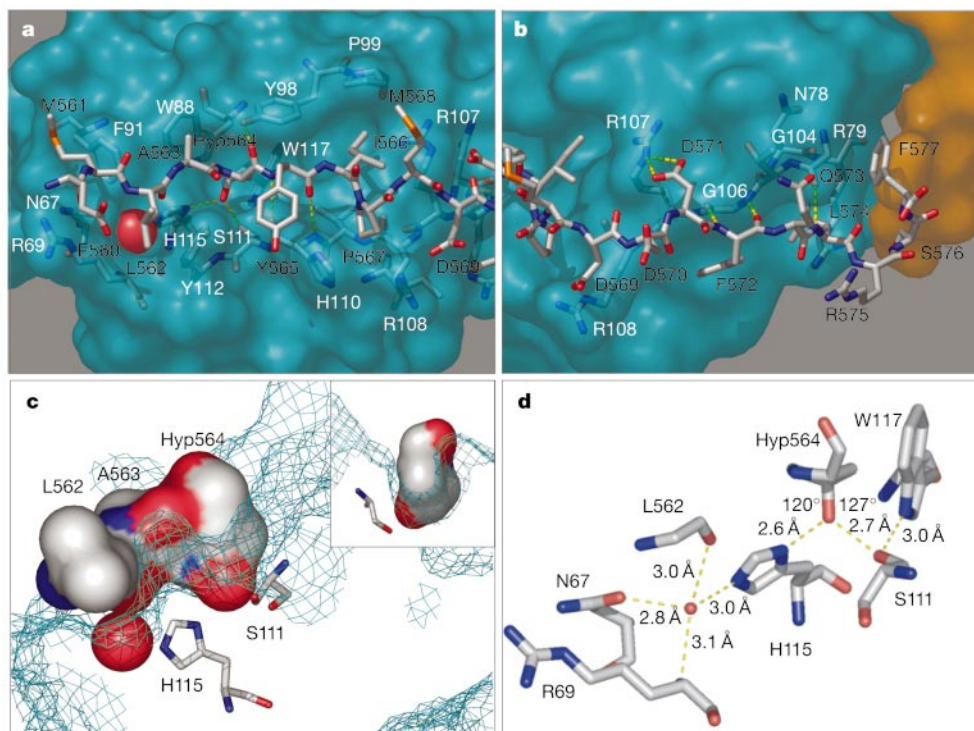


Figure 3 CODD-pVHL interactions. **a, b**, Views of site 1 (**a**) and site 2 (**b**). CODD and CODD-contacting residues of pVHL are shown as stick models (colour scheme as in Fig. 1), whereas pVHL, elongin C and elongin B are rendered as van der Waals surfaces using the colour scheme of Fig. 2. Dotted lines represent electrostatic bonds. Residues for pVHL are labelled in white; those for CODD are in black. **c**, Views of the Hyp-binding pocket, illustrating its surface complementarity. The van der Waals surface of the ODD motif, LAHyp, is rendered solid and coloured by atom type (as in Fig. 1). The surface of pVHL is

rendered as a mesh object with both the front and rear planes clipped away. Inset, side view of Hyp. **d**, The hydrogen-bonding network involved in binding of the Hyp564 hydroxyl group. The side chain of L562 (CODD) has been omitted for clarity. The red sphere in **a, c** and **d** represents a key water molecule. **a** has an approximately similar orientation to that of Fig. 2a, whereas the view in **b** is rotated from that in **a** by 70 degrees about the vertical axis.

to hypoxia, including angiogenesis and metabolic adaptation¹⁹, has provided a new impetus for the therapy of ischaemic diseases through the activation of HIF. Transgenic expression of stabilized HIF-1 α molecules that lack the pVHL recognition motifs leads to hypervascularity without leakage or inflammation²⁰, indicating that proteolytic escape of HIF-1 α can drive a productive angiogenic response. The specificity intrinsic to the rigid hydroxyproline-binding pocket of pVHL revealed here suggests that it could be targeted for development of small molecules to block HIF-1 α recognition by pVHL. □

Methods

Peptide synthesis

Biotinylated and unmodified human CODD peptides (residues 549–582) were synthesized by Genemed Synthesis. The other peptides were synthesized by Biopeptide. Peptide purity was assessed by mass spectrometry and high-performance liquid chromatography.

Protein expression, purification and crystallization

The proteins pVHL, elongin B and elongin C were co-expressed in *Escherichia coli* using vectors provided by N. Pavletich. The protein complex was purified essentially as described previously¹⁵. Purified VCB was mixed with excess biotinylated and hydroxylated CODD peptide, and was further purified on a Superdex 200 size-exclusion column (Pharmacia) in buffer containing 5 mM HEPES (pH 7.4), 200 mM NaCl and 1 mM dithiothreitol. Crystallization of the heterotetrameric complex (at 8 mg ml⁻¹) was carried out by sitting-drop vapour diffusion at 22 °C. The mother liquor consisted of 100 mM HEPES (pH 7.4), 40% polyethylene glycol 2000 monomethylester and 200 mM (NH₄)₂SO₄.

Crystallographic data collection and structure refinement

Two native data sets were collected. The crystals were cooled in their mother liquor to 100 K. Crystallographic data were indexed and scaled using the HKL2000 package²¹. The crystals belonged to the space group *P*₄₃₂ with one complex per asymmetric unit. A data set, to 3.2 Å resolution, was collected with a Rigaku X-ray generator (100 mA and 50 kV) and a MAR345 imaging plate detector. Coordinates for the VCB complex (Protein Data Bank code 1VCB) yielded a molecular replacement solution (program EPMR²²). A difference map revealed the peptide. Data to 2.0 Å resolution were subsequently collected on beamline ID14-2 at the European Synchrotron Radiation Facility using an ADSC Quantum 4 detector. Cycles of rebuilding and refinement were carried out using the programs O²³ and CNS²⁴. The current *R*_{cryst} and *R*_{free} are 22.6% and 27.8%, respectively (r.m.s. deviations from ideal bond lengths and angles are 0.006 Å and 1.3°, respectively). Other statistics are listed in Supplementary Information. The following residues are missing in the current model: 105–118 in elongin B, 50–57 in elongin C, 54–60 and 211–213 in pVHL, 549–559 and 578–582 in CODD. The backbone geometry of the refined model is satisfactory, with 89.3% of the residues in favourable conformations (program PROCHECK²⁵). Two residues, H10 and D47 of elongin B, are in the disallowed region. The electron density around these two residues is excellent and their unusual main-chain geometry is due to their structural role in stabilizing two β -turns.

Structural analyses and graphics presentation

Solvent-accessible surface areas were calculated by NACCESS²⁶. Least-squares structural alignment was performed with LSQMAN²⁷. All structural graphic presentations were rendered with PyMOL (<http://pymol.sourceforge.net>). Intermolecular van der Waals contacts and hydrogen bonds were analysed by CONTACT, and surface complementarity was calculated by Sc¹⁶ (a score of 1 is assigned to perfect complementarity, and 0 to none); both programs are implemented in CCP4²⁸.

Surface plasmon resonance

Experiments were performed in duplicate at 25 °C on a BIAcore 2000 using streptavidin-coated sensor chips and analysed using the BIAevaluation software (version 3.0) (BIAcore AB). The data were collected at a rate of 10 Hz, and all analyses were performed with VCB complexes injected over biotinylated peptides immobilized on the sensor chips. Peptides for site 1 were synthesized with an extra N-terminal cysteine residue to which biotin tags were chemically attached using a biotin-PEO-maleimide linker (which contains a spacer arm of ~29 Å; Perbio Science), whereas peptides containing both sites 1 and 2 were synthesized with biotin attached directly at the N terminus. The concentration of the VCB complex was estimated by absorbance at 280 nm, using the theoretical extinction coefficient 24,540 M⁻¹ cm⁻¹. Kinetics experiments used a high flow rate (100 μ l min⁻¹) and a low surface density of ligand to minimize mass transport limitation²⁹. Each experiment was performed with five dilutions of the VCB complex: 4.88, 2.44, 1.22, 0.610 and 0.305 nM. Association and dissociation constants were obtained by global fitting of the data to a 1:1 Langmuir binding model. For the competition experiments, the VCB complex was pre-incubated briefly at room temperature with various concentrations of nonbiotinylated peptides. The relative rate of binding estimated from the initial slope of the binding curves was used as a measure of inhibition. For the steady-state binding experiment, responses of VCB binding to immobilized, unhydroxylated CODD were subtracted from binding to a blank cell over a range of concentrations.

Received 10 April; accepted 15 May 2002; doi:10.1038/nature00767.

Published online 5 June 2002.

- Wang, G. L., Jiang, B. H., Rue, E. A. & Semenza, G. L. Hypoxia-inducible factor 1 is a basic-helix-loop-helix-PAS heterodimer regulated by cellular O₂ tension. *Proc. Natl Acad. Sci. USA* **92**, 5510–5514 (1995).
- Maxwell, P. H. *et al.* The tumour suppressor protein VHL targets hypoxia-inducible factors for oxygen-dependent proteolysis. *Nature* **399**, 271–275 (1999).
- Ohh, M. *et al.* Ubiquitination of hypoxia-inducible factor requires direct binding to the β -domain of the von Hippel-Lindau protein. *Nature Cell Biol.* **2**, 423–427 (2000).
- Ivan, M. *et al.* HIF α targeted for VHL-mediated destruction by proline hydroxylation: implications for O₂ sensing. *Science* **292**, 464–468 (2001).
- Jaakkola, P. *et al.* Targeting of HIF- α to the von Hippel-Lindau ubiquitylation complex by O₂-regulated prolyl hydroxylation. *Science* **292**, 468–472 (2001).
- Yu, F., White, S. B., Zhao, Q. & Lee, F. S. HIF-1 α binding to VHL is regulated by stimulus-sensitive proline hydroxylation. *Proc. Natl Acad. Sci. USA* **98**, 9630–9635 (2001).
- Masson, N., Willam, C., Maxwell, P. H., Pugh, C. W. & Ratcliffe, P. J. Independent function of two destruction domains in hypoxia-inducible factor- α chains activated by prolyl hydroxylation. *EMBO J.* **20**, 5197–5206 (2001).
- Hershko, A. & Ciechanover, A. The ubiquitin system. *Annu. Rev. Biochem.* **67**, 425–479 (1998).
- Tyers, M. & Jorgensen, P. Proteolysis and the cell cycle: with this RING I do thee destroy. *Curr. Opin. Genet. Dev.* **10**, 54–64 (2000).
- Kamura, T. *et al.* Rbx1, a component of the VHL tumour suppressor complex and SCF ubiquitin ligase. *Science* **284**, 657–661 (1999).
- Bruick, R. K. & McKnight, S. L. A conserved family of prolyl-4-hydroxylases that modify HIF. *Science* **294**, 1337–1340 (2001).
- Epstein, A. C. *et al.* *C. elegans* EGL-9 and mammalian homologs define a family of dioxygenases that regulate HIF by prolyl hydroxylation. *Cell* **107**, 43–54 (2001).
- Pugh, C. W., O'Rourke, J. F., Nagao, M., Gladde, J. M. & Ratcliffe, P. J. Activation of hypoxia-inducible factor-1; definition of regulatory domains within the α subunit. *J. Biol. Chem.* **272**, 11205–11214 (1997).
- Huang, L. E., Gu, J., Schau, M. & Bunn, H. F. Regulation of hypoxia-inducible factor 1 α is mediated by an O₂-dependent degradation domain via the ubiquitin-proteasome pathway. *Proc. Natl Acad. Sci. USA* **95**, 7987–7992 (1998).
- Stebbins, C. E., Kaelin, W. G. Jr & Pavletich, N. P. Structure of the VHL–ElonginC–ElonginB complex: implications for VHL tumour suppressor function. *Science* **284**, 455–461 (1999).
- Lawrence, M. C. & Colman, P. M. Shape complementarity at protein/protein interfaces. *J. Mol. Biol.* **234**, 946–950 (1993).
- Beroud, C. *et al.* Software and database for the analysis of mutations in the VHL gene. *Nucleic Acids Res.* **26**, 256–258 (1998).
- Fersht, A. R. *et al.* Hydrogen bonding and biological specificity analysed by protein engineering. *Nature* **314**, 235–238 (1985).
- Semenza, G. L. Regulation of mammalian O₂ homeostasis by hypoxia-inducible factor 1. *Annu. Rev. Cell Dev. Biol.* **15**, 551–578 (1999).
- Elson, D. A. *et al.* Induction of hypervascularity without leakage or inflammation in transgenic mice overexpressing hypoxia-inducible factor-1 α . *Genes Dev.* **15**, 2520–2532 (2001).
- Otwinski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307–326.
- Kissinger, C. R., Gehlhaar, D. K. & Fogel, D. B. Rapid automated molecular replacement by evolutionary search. *Acta Crystallogr. D* **55**, 484–491 (1999).
- Jones, T. A., Zou, J. Y., Cowan, S. W. & Kjeldgaard, M. Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr. A* **47**, 110–119 (1991).
- Brunger, A. T. *et al.* Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr. D* **54**, 905–921 (1998).
- Laskowski, R. A., Rullmann, J. A., MacArthur, M. W., Kaptein, R. & Thornton, J. M. AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR. *J. Biomol. NMR* **8**, 477–486 (1996).
- Hubbard, S. J. & Thornton, J. M. 'NACCESS', *Computer Program* Version 2.1.1 (Department of Biochemistry and Molecular Biology, Univ. Coll. London, London, 1996).
- Kleywegt, G. J. Use of non-crystallographic symmetry in protein structure refinement. *Acta Crystallogr. D* **52**, 842–857 (1996).
- CCP4 The CCP4 suite: Programs for protein crystallography. *Acta Crystallogr. D* **50**, 760–763 (1994).
- Myszka, D. G. Kinetic analysis of macromolecular interactions using surface plasmon resonance biosensors. *Curr. Opin. Biotechnol.* **8**, 50–57 (1997).

Supplementary Information accompanies the paper on Nature's website (<http://www.nature.com/nature>).

Acknowledgements

We thank the staff at beamline ID14-2 of the ESRF, France. We are grateful to A. van der Merwe and A. Kearney for advice and assistance with the surface plasmon resonance experiments. L. M. McNeil assisted with the NMR analysis. This work was supported by Cancer Research UK and the Wellcome Trust. W.-C.H. and M.I.W. were recipients of a Human Frontiers long-term fellowship and a Natural Sciences and Engineering Research Council of Canada postdoctoral fellowship, respectively. The Medical Research Council supports C.W.P. and D.I.S., and Cancer Research UK supports E.Y.J.

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to E.Y.J. (e-mail: yvonne@strubi.ox.ac.uk) or P.J.R. (e-mail: peter.ratcliffe@ndm.ox.ac.uk). Coordinates and structure factors have been deposited in the Protein Data Bank under accession codes 1LQB and RCSB016181.

Detection, DNA, diagnostics

A selection of recent introductions including apoptosis detection kits.

Selective β_3 adrenoceptor antagonist

Sigma-Aldrich www.sigma-aldrich.com

Mass observation

Sigma-Aldrich describes this as the first commercially available selective β_3 -adrenoceptor antagonist for use in the study of obesity. β -Adrenoceptors belong to the family of G-protein-coupled receptors and are widely distributed throughout central and peripheral tissues. The three known subtypes are designated β_1 , β_2 and β_3 . The β_3 receptor subtype is the predominant type found in adipose tissue, and it has been shown that activation of these receptors stimulates lipolysis in white adipose tissue, while increasing energy expenditure in brown adipose tissue. β_3 -adrenoceptor-mediated responses are of interest to researchers investigating the role of these receptors in mechanisms underlying the regulation of body weight.

Eclipse DNA detection kits

Biotest www.biotestusa.com

Competition is in the blood

Eclipse CMV (cytomegalovirus) and EBV (Epstein-Barr virus) detection kits from Biotest provide quantitative detection of human CMV or EBV DNA extracted from whole blood. Using PCR methodology, the kits exploit simultaneous competitive PCR amplification and competitive hybridization of target DNA and an internal competitive standard (ICS) DNA in order to determine viral load. Competitive amplification and hybridization ensure that the ICS DNA

competes with the target DNA during each step of the assay, so that, throughout the process, any variable has the same effect on both the ICS and the target DNA. Using a four-point reference line, the unknown CMV or EBV DNA copy number can be quantified by means of automated data analysis. The kits are supplied with enough reagent to perform up to 24 tests, two pre-coated 96-well microwell capture plates, virus-specific amplification primer sets, ICS detection probes, positive and negative controls and all reagents.

Apo E2, E3 and E4

Biodesign www.biodesign.com

Three novel recombinant apolipoproteins

Biodesign has introduced three recombinant apolipoproteins, E2, E3 and E4. Apolipoprotein E (Apo E) is a ligand for the low-density lipoprotein (LDL) receptor and may play an important role in nerve degeneration and regeneration. It is studied in association with neuron function and disorders such as Alzheimer's disease and coronary heart disease. Each Apo E isoform retains full biological activity, enabling researchers to study the interactions of Apo E with other proteins such as β -amyloid, tau and the LDL receptor. The apolipoproteins are produced by baculovirus-mediated expression in *Spodoptera frugiperda* cells and expressed and purified (by affinity chromatography) in their native conformation. Each is supplied in 0.7 M ammonium bicarbonate buffer in a 50 μ g vial.

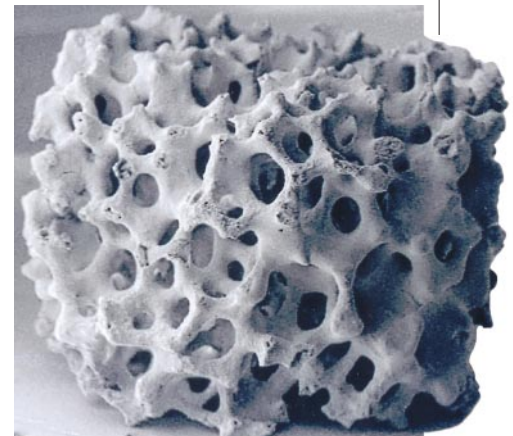
Western Lightning/DNA Thunder

PerkinElmer Life Sciences

www.perkinelmer.com

Protein and nucleic acid detection

Western Lightning and DNA Thunder are chemiluminescence reagents that can be used in detecting many different blot-based protein and nucleic acid studies, including western blotting assays, non-radioactive nucleic acid labelling and DNA slot blot applications. They comprise a family of kits with both high sensitivity and conventional strength reagents for horseradish peroxidase (HRP)-based assays, as well as kits designed for alkaline phosphatase (AP)-based detection assays in each family. Both families feature high sensitivity detection, up to four times higher than other brands available to researchers. Neither product family requires a change in protocol, making it easy for researchers to integrate these reagents into their western blotting or nucleic acid detection protocols.



Bone idol: BD Biosciences' scaffolding.

3D Calcium Phosphate Scaffold

BD Biosciences www.bdbiosciences.com

Bone up on the essentials

Fashioned out of mineralized calcium phosphate bioceramic, BD's scaffold is designed for use in *in vitro* and *in vivo* analysis of bone metabolism and *in vivo* analyses of cartilage regeneration. The three-dimensional structure mimics the *in vivo* environment to induce tissue formation or promote tissue repair. Bone metabolism is a complex process that involves the resorption of existing bone by osteoclasts and the subsequent formation of a new bone matrix by osteoblasts. These activities are essential for bone remodelling, regeneration and repair. For tissue engineering applications, bone biomaterials must be capable of supporting the functional properties of osteogenic cells. While bone cells are unable to remodel materials such as hydroxyapatite, studies have shown that the mineralized calcium phosphate substrate supports bone remodelling *in vivo*. An optimized assay and protocol are included.

VasoTACS

Trevigen

www.trevigen.com

New kit in circulation

The VasoTACS *in situ* apoptosis detection kit was developed by researchers interested in angiogenesis and vascular disorders. The kit detects DNA fragmentation using TdT, special enzymes, and counterstains tailored for use in vascular tissue. It comes supplied with permeabilization and labelling reagents, nuclease to generate a positive control with the user's own samples, and TACS Blue Label detection reagent to provide contrast for viewing of labelled cells against the Red Counterstain C.



Biotest: CMV and EBV detection.



D light: a new assay from IDS.

Octeia

Immunodiagnostic Systems www.idsltd.com

A simple ELISA for vitamin D

Octeia is a 25-hydroxy vitamin D ELISA for the quantitative determination of 25-hydroxy vitamin D in human serum or plasma. It is described as a much-simplified method of determining patient vitamin D status. Serum or plasma samples are used directly in the kit, with no sample extraction necessary. Sensitivity is $<2.4 \text{ ng ml}^{-1}$ ($<6 \text{ nmol l}^{-1}$), with a working range of $2.4\text{--}144 \text{ ng ml}^{-1}$ ($6\text{--}30 \text{ nmol l}^{-1}$). The assay is available in a 96-well format and is suitable for automation on existing automated ELISA platforms.

Insite

National Diagnostics

www.nationaldiagnostics.com

Bright idea

Insite is a fluorescent in-gel stain for SDS-PAGE applications. It incorporates a dye/buffer combination that stains proteins during the gel run. Following electrophoresis, bands containing as little as 10 ng of protein can be detected. The Insite kit contains buffer and dye concentrates that are combined to formulate the cathode chamber buffer. The anode chamber is filled with tris-glycine-SDS buffer and the gel run following standard SDS-PAGE procedures. After the run, the gel is destained for 30 minutes in deionized water and then visualized by ultra-violet transillumination. The protein bands fluoresce yellow-orange.

Alpha-1-antichymotrypsin antibody

BioGenex www.biogenex.com

For the detection of liver and other cancers

This monoclonal antibody, produced by Clone 7B2, specifically recognizes human α -1-antichymotrypsin. Immunohistochemical staining using this antibody shows strong cytoplasmic staining in healthy liver cells, but no staining or very weak staining in hepatoma.

The sensitivity of the antibody to malignant cells may help to detect primary liver cancers and other cancers, such as those of the thyroid. BioGenex offers the antibody in a concentrated form or as a ready-to-use formulation.

Oestrogen receptor β antibody

Serotec

www.serotec.co.uk

They've made a concentrated effort

Serotec's oestrogen receptor β (Er β) antibody is now available in a concentrated format. The $10\times$ concentrate is especially suitable for rapid staining procedures. Two oestrogen receptors have been described, Er α and Er β . These cytoplasmic proteins bind oestrogens and migrate to the nucleus, where they regulate DNA transcription. Sites of expression of the receptors in the ovary and testis are essential to elucidate the role of oestrogens in ovarian and testicular physiology. Er α has also been found to be expressed in breast cancer tumours, either alone or in combination with Er β . Tumours co-expressing both receptors were node-positive and tended to be of a higher grade. Thus, Er β may prove to be a prognostic marker in breast cancer.

DIASTAT

Axis-Shield

www.axis-shield.com

Upgrades, enhancements, and a new range of slides

DIASTAT enzyme immunoassays are designed for the differential diagnosis of autoimmune diseases such as systemic lupus erythematosus, Sjögren's syndrome, undifferentiated connective tissue disease, polymyositis and scleroderma. The upgraded DIASTAT Total Cardiolipin kit features a new synthetic cardiolipin antigen and a more robust inner compartment. Enhancements to the DIASTAT Anti-TPO assay include the use of human recombinant TPO antigen, lower clinical cut-off reflecting increased sensitivity and specificity, and a narrower measuring range. DIASTAT IFA slides include monkey oesophagus for skin and endomysial antibodies, human PMN cells for ANCA screening, *Crithidia luciliae*

for dsDNA antibody configuration, human epithelial cells for ANA detection and rat liver/kidney/stomach for general autoantibody screening. All products are suitable for manual and automated ELISA applications.

Universal ISH kits

InnoGenex

www.innogenex.com

Better for all

In situ hybridization (ISH) kits from InnoGenex contain non-radioactive systems for the localization of DNA and RNA sequences in all cells and tissues, including those of rodent species. The kits are available in two formats suitable for the detection of either fluorescein- or digoxin-labelled probes. Each kit contains sufficient reagents to stain 60 slides, and comprises proteinase K solution, RNase block, hybridization buffer, protein block, biotinylated anti-fluorescein or anti-digoxin antibody, enzyme-streptavidin or fluorochrome conjugate, activation buffer (for phosphatase kits), chromogen/substrate, counterstain, mounting medium and phosphate buffered saline.

omniflex LT

Bruker Daltonics

www.daltonics.bruker.com

Simple mass spectrometry for MALDI-TOF

Priced at \$89,900 in the US, the omniflex LT is aimed at bringing affordability for a high-quality, automated large biomolecular mass reader, making MALDI-TOF accessible to many clinical, biology and chemistry laboratories. The omniflex LT is an automated bench-top MALDI-TOF designed for facile mass analysis by users not experienced in mass spectroscopy. Based on the proven omniflex platform, the LT is supported by a variety of MALDI-TOF accessories and services available from Bruker Daltonics, including AnchorChips, purified matrices, mass calibrants, training tools, assays and bioinformatics software.

These notes are compiled in the Nature office from information provided by the manufacturers.

Contacts

Publisher: Ben Crowe

Editor: Paul Smaglik

Marketing Manager: David Bowen

European Head Office, London

The Macmillan Building
4 Crinan Street
London N1 9XW, UK
Tel +44 (0) 20 7843 4961
Fax +44 (0) 20 7843 4996
e-mail: naturejobs@nature.com

Senior European Sales Manager:
Nevin Bayoumi (4978)

UK/ RoW/ Ireland:

Matt Powell (4953)
Andy Douglas (4975)
Frank Phelan (4944)
Amelie Pequignot (4974)
Laura Pearson (4977)

Netherlands/ Italy/ Iberia:

Evelina Rubio Hakansson (4973)
Scandinavia: Silje Opstrup (4994)

Production Manager:

Billie Franklin
To send materials use London
address above.
Tel +44 (0) 20 7843 4814
Fax +44 (0) 20 7843 4996
e-mail: naturejobs@nature.com

International

Advertising Coordinator:

Hind Berrada (4935)

Naturejobs web development:

Tom Hancock

Naturejobs online production:

Ben Lund

European Satellite Offices

France/ Belgium:

Christine Niox-Chateau
Tel + 33 (0) 1 43 20 16 51
Fax + 33 (0) 1 43 20 51 52
e-mail: c.nioxchateau@nature.com

Germany/ Austria/ Switzerland:

Patrick Phelan, Odo Wulffen
Tel + 49 89 54 90 57 11/-2
Fax + 49 89 54 90 57 20
e-mail: p.phelan@nature.com
o.wulffen@nature.com

US Head Office, New York

345 Park Avenue South,
10th Floor, New York, NY 10010-1707
Tel +1 800 989 7718
Fax +1 800 989 7103
e-mail: naturejobs@natureny.com

US Sales Manager:

Peyton Mason

US Advertising Coordinator:

Ashly de Leon

Japan Head Office, Tokyo

MG Ichigaya Building (5F),
19-1 Harakata-machi,
Shinjuku-ku,
Tokyo 162-0841
Tel +81 3 3267 8751
Fax +81 3 3267 8746
e-mail: k.johnson@naturejp.com

Asia-Pacific Advertising Manager:

Kevyn Johnson

naturejobs

The digital revolution

Classes in computational biology are among the most popular offerings of the advanced-course season. But they aren't the only show in town. For example, one of Cold Spring Harbor Laboratory's stalwarts, a class on mouse embryology, celebrated its twentieth anniversary this month. And it still draws four or five times as many applicants as it can accommodate, even though — or perhaps because — “the course in no way bears any resemblance to what was taught 20 years ago”, says Dave Stewart, director of meetings and courses at the laboratory.

Change, when it comes to course curricula, does scientists good. And there definitely is change afoot, with computational components taking up larger portions of the curriculum for subjects such as proteomics, genetics and gene expression. This allows more discipline-specific techniques to be disseminated — such as methods of interpreting microarray data, or using computers to analyse different kinds of neuroscience imaging data. Of course, these tailored courses complement the more general bioinformatics offerings.

Interest in both tailored and general bioinformatics training shows no sign of waning. A bioinformatics course sponsored by the French research agency INSERM is already oversubscribed with just under a month to go before the closing date for applications. In addition, almost all of the UK Wellcome Trust's genomics and proteomics courses feature computational components, and informatics is a growing aspect in many of Cold Spring Harbor Laboratory's courses — but not, of course, the mouse-embryology one.

So how can the enduring appeal of the mouse course be explained in these high-tech times? Stewart believes he knows: “A lot of people are still interested in learning from regular biology.”

Paul Smaglik

Naturejobs editor



Contents

CAREERS & RECRUITMENT

Cell biology gets computerized p4

Business opportunities in a virtual world p6

Mathematical models give Japan a fresh approach p7

WWW.NATUREJOBS.COM

Career centre
Information on the scientific job market

FOCUS

SPOTLIGHT

RECRUITMENT

SCIENTIFIC ANNOUNCEMENTS

SCIENTIFIC EVENTS

Silicon dreams in the biology lab

For many biologists, the idea of creating a computer model of a living cell is anathema. But for mathematicians and physicists, the pursuit of such a goal is proving irresistible, says Diane Gershon.

Computational cell biology, although not a new field, has not been given much weight until relatively recently. But as genomic and proteomic data grow more prevalent and computers become more powerful, scientists are making greater efforts to build computational models for cellular activity — from whole cells to entire systems, complete with gene expression, and regulatory and metabolic pathways. Several fledgling institutions and initiatives mean more opportunities for biologists willing to marry their skills with maths and computer science. But, ironically, biologists seem to be the most resistant to this new area of activity.

ENTRENCHED RESISTANCE

John Carson, a professor in the biochemistry department at the University of Connecticut Health Center in Farmington, says that he has found it easy to recruit mathematicians, physicists and engineers to work on computational cell biology because it is a grand and novel challenge for them. “We’ve found it a little bit more difficult to recruit biologists, or to convince biologists that computational approaches are

going to be valuable,” he says, “because many biologists feel that biology is just too complicated to deal with computationally.”

At least some of biologists’ resistance lies in their training. For the new work, they need to understand differential equations and rate constants, for example, and historically biology has not been taught that way. In the experience of John Tyson, a professor in the department of biology at Virginia Tech in Blacksburg, most of the people who move into computational biology do not come from traditional biology backgrounds but rather

from physical chemistry, chemical engineering, applied mathematics, numerical analysis and biophysics.

More resistance lies in taking a leap of faith away from their tried-and-tested methods. Many mainstream cell and molecular biologists still need convincing that it can be done, says Leslie Loew, director of the Center for Biomedical Imaging Technology at the University of Connecticut Health Center. “There are a lot of biologists who are very content to chip away at their own very highly restricted problem and not really try to build a full picture of cell-biological systems,” he says.

One criticism levelled at computational cell biology in the past was that it was often pretty irrelevant to what was going on in the lab. Loew agrees that this has not helped its image. “But if computational modelling is closely linked to experiment,” he says, “then it becomes a very valuable adjunct to experimentation.”

BEYOND BIOINFORMATICS

Modelling is a logical next step from existing bioinformatics, argues Tyson. “People with a little further vision are seeing that computational cell biology is just around the corner,” he says.

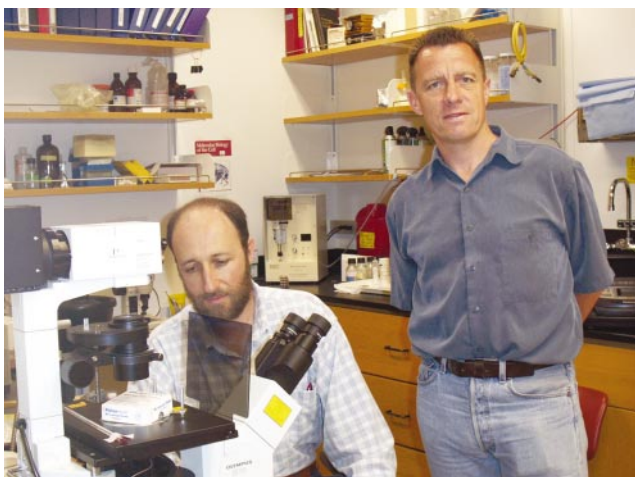
Identifying genes, determining the structure of proteins and investigating protein–protein interactions is all very well but “we think that gives you a kind of static picture of the cell,” Tyson says. To understand how the components of a cell interact in a dynamically changing environment, “you need to convert the interactions into mathematical equations and animate them on a computer,” he explains.

Some proponents of computational biology think that an increase in the number of initiatives that combine experimentation with modelling might bridge the gap and quash the scepticism. Such efforts would require experimental biologists to rub shoulders with theoretical physicists, mathematicians, engineers and computer scientists.

Several recent initiatives — although broader in scope than just computational cell biology — are already set up this way, such as the Institute for

John Tyson sees computational cell biology as the logical extension of bioinformatics.





Systems Biology in Seattle, Washington, which is headed by Leroy Hood. Another example is the Institute for Quantitative Biomedical Research, a multi-campus effort, based at the University of California, San Francisco, and directed by Marvin Cassman, former director of the National Institute of General Medical Sciences.

On a more modest scale, the University of California, Davis, will this autumn open a new Laboratory of Cell and Computational Biology within the Center for Genetics and Development, which will house both experimental and theoretical scientists under one roof. The lab will focus on a combination of experimental cell biology, under the direction of Jonathan Scholey of the university's molecular and cell-biology section, and mathematical modelling of cell-biological phenomena, under the direction of Alex Mogilner of the institute's department of mathematics. A particular emphasis for the new lab will be on cytoskeleton dynamics, cell motility, mitosis and cell division.

IMPROVING ACCESSIBILITY

These types of initiative are providing a few examples of high-profile, but relatively isolated, activity. If computational cell biology is to be accepted by more mainstream molecular and cell biologists, "we have to have some user-friendly interfaces that lower the activation barrier for experimental biologists to start thinking about mathematical models", says Tyson.

Efforts such as the Virtual Cell project, developed as part of the University of Connecticut's National Resource for Cell Analysis and Modeling (NRCAM) and funded by the National Institutes of Health (NIH), have helped. Virtual Cell, which is available free on the Internet for non-commercial use, provides a generalized modelling and simulation framework that biologists without a mathematical background — and with a little training — can use.

Other examples include the E-CELL simulation software, developed at Keio University in Japan (see "Computerized role models", page 7), and the Gepasi biochemical kinetics simulator, developed at the University of Wales, Aberystwyth, UK.

The US defence department's Defense Advanced Research Projects Agency also began a major biocomputing initiative last year. The agency plans to

spend US\$50 million over five years to fund projects to create an open-source framework and tool kit for modelling dynamic cellular-network functions.

GETTING THE MESSAGE OUT

The field is gaining visibility. "Whereas before we mostly talked among ourselves, now we're trying as hard as we can to branch out and take the message to practising experimental cell biologists," says Tyson.

The second International Symposium on Computational Cell Biology, organized by the NRCAM, will be held on 22–25 March 2003 in Lenox, Massachusetts. Last year's meeting attracted a healthy mix of theoreticians and experimentalists, although many of the latter were already converts. Al Gilman, Nobel laureate and professor of pharmacology at the University of Texas Southwestern Medical Center in Dallas, will be the keynote speaker at next year's meeting.

Gilman heads the Alliance for Cellular Signaling — a consortium funded in part by one of the NIH's 'glue' grants — which is studying the G-protein-mediated, and related, signalling systems in two different cell types (see *Naturejobs* 4–5; 19 April 2001). The initiative has had a computational modelling component from the outset, albeit a smaller part of the overall project. Another glue grant is focused on the area of cell migration.

But how well and how fast computational cell biology meshes with the mainstream remains an open question. The development of interdisciplinary research and training initiatives and user-friendly computational modelling tools that bring quantitative modelling methods to the general biologist will undoubtedly help. When asked: "What will it take?" Tyson quips: "Some very striking and counterintuitive prediction that's confirmed by experiment." And perhaps more biologists willing to take a leap out of their experimentally driven boxes. ■

Diane Gershon is assistant editor, new technology, at *Nature Medicine*.

Virtual Cell modelling and simulation framework

♦ www.nrcam.uchc.edu

Gepasi biochemical kinetics simulation package

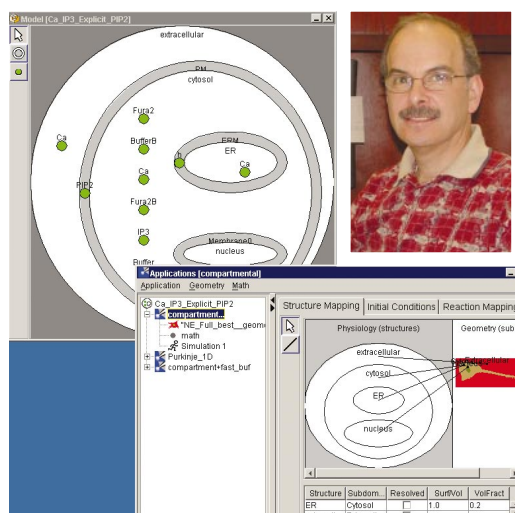
♦ www.gepasi.org

Alliance for Cellular Signaling

♦ www.cellularsignaling.org

Cell Migration Consortium

♦ www.cellmigration.org



Leslie Loew believes computational efforts such as the Virtual Cell project have a lot to offer experimental biologists.

Cell biology gets plugged in

The choice of potential drug targets thrown up by genomics data is overwhelming, which is why several firms are now offering drug companies a model solution. Diane Gershon reports.

A handful of companies have surfaced in recent years that are pinning their hopes on computational cell biology. Their rationale is simple. Many drug companies have invested heavily in bioinformatics to deal with the deluge of data flowing from the Human Genome Project, as well as from proteomics and DNA-chip technologies. Although this has produced many more potential drug targets, it has not necessarily led to a better understanding of the clinical relevance of the targets or of drugs directed against them. Given the high cost of clinical trials, tools that help drug firms to decide where to focus their research and resources could potentially be of considerable value to them.

With this in mind, start-ups such as Entelos in Menlo Park, California; Gene Network Sciences (GNS) in Ithaca, New York; Physiome Sciences in Princeton, New Jersey; Cellnomica in Fort Myers, Florida; and Genomatica in San Diego, California, are combining mathematical theory and computer simulation with experimental data to develop computational modelling technologies which, they hope, will turn data into real predictions about drug targets and candidates.

"The motivation comes from the clinical-trial end of drug development, but the real source of the issue is down to target selection," says Thomas Paterson, chief scientific officer at Entelos. The company's roots go

back to the early 1990s, when several of the company's founders began to apply simulation and decision-support technologies, similar to those used in the aerospace and defence industries, to problems in drug research. This became the company's core PhysioLab technology, which it now uses to develop dynamic, large-scale computer models of human disease. It specifically focuses on chronic, complex diseases such as obesity and diabetes.

Rather than modelling cells, "we actually model multiple organ, tissue and cell systems that participate in a disease process", says Paterson. "What models allow you to do is to be very explicit with your hypothesis, and then to test the implications of that hypothesis in a very rigorous, systematic way." But it is results that count, and Paterson points to several case studies where, he says, Entelos successfully predicted the outcome of two different asthma clinical trials.

BACKGROUND COUNTS

But modellers are hard to come by, says Colin Hill, chief executive and co-founder of GNS, which is focusing its efforts on oncology and infectious diseases. These are individuals with a background in engineering or physics, which provides familiarity with mathematical equations and numerical analysis, but also with some proficiency in computer programming and some knowledge of biology. "Very few people have these three backgrounds," he says. It takes someone who goes well beyond bioinformatics and the creation of data-mining algorithms and sequence-search algorithms, he adds.

Earlier this month, at the Beyond Genome conference in San Diego, GNS announced that it had created the largest known data-driven computer model of a human cancer cell. But Hill acknowledges this is only a first step and that the model will be refined. "It still only represents a couple of per cent of the circuitry within the cell," he says.

At the moment, a lack of quantitative time-series data is limiting the field, says Hill. Although GNS has its own wet lab to collect some of this data, it has hooked up with some proteomics firms and is seeking other partners on the experimental side. And, even with its 200 processors from IBM, Hill says GNS will very quickly need more computing power. **Diane Gershon**

Gene Network Sciences ♦ www.gnsbiotech.com

Entelos ♦ www.entelos.com

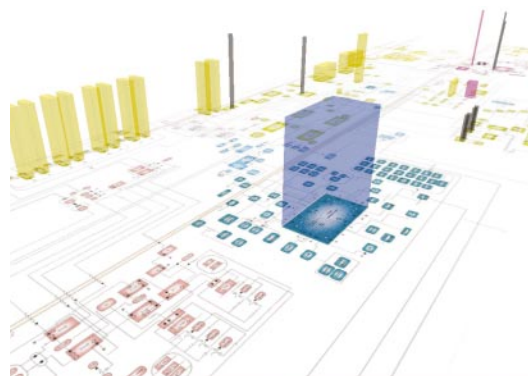
Physiome Sciences ♦ www.physiome.com

Genomatica ♦ www.genomatica.com

Cellnomica ♦ www.cellnomica.com



Colin Hill sees huge potential for model cells (below).



Computerized role models

Japan's push to create a virtual cell signals a new approach to research, says Robert Triendl.

IAB

In many ways, the year-old Institute for Advanced Biosciences at Keio University in Tsuruoka is an anomaly of Japanese science. It is run by a relatively young scientist, uses short-term appointments, and it gives its junior faculty members a great deal of academic freedom.

The institute was established last year with a grant from the Yamagata prefecture with additional funding from Keio University, Japan's top-ranked private university. It is built around E-CELL, an ambitious idea by its director, Masaru Tomita, to develop a realistic computer simulation of an entire living cell.

Tomita trained as a computer scientist at Keio and at Carnegie Mellon University in Pittsburgh, where he also worked for several years as an assistant professor. After he returned to Japan in the early 1990s, his group at Keio's Fujisawa campus, along with scientists at The Institute of Genomic Research in Rockville, Maryland, completed the first 'virtual cell' in 1997, based on the workings of 127 'virtual' genes selected from the genome of the bacterium *Mycoplasma genitalium*.

Tomita quickly recognized that to build better models, he needed more data. But as such data were not available, there was little he could do except build the data sets on his own. His chance came when the Yamagata prefecture approached Keio University with the proposal to set up a campus in Tsuruoka.

DATA HUNTING

At the Institute for Advanced Biosciences, Tomita has assembled a research team that covers a broad range of specialisms, including enzyme engineering, analytical chemistry, genetic engineering, computer sciences and mathematics. Although most of the scientists have a background in molecular biology or biochemistry, work at the institute is orientated towards creating realistic computer models of living cells.

Today, there are about 30 scientists and roughly the same number of graduate students working at the institute. Most of the staff scientists are younger researchers in assistant-professor positions. All appointments, including Tomita's own as director,



are renewable after three years.

Although this policy is still unusual in a country with an extremely tight academic labour market, Tomita says he voted for it despite the university's reservations. "Three years is a short period of time, but I think it's a good incentive," he says. "More importantly, a limited-term contract provides young scientists with much more freedom than a tenured faculty position. There is nobody who tries to control you." Nevertheless, many of the senior staff hold joint appointments at other parts of the university.

Already, most of the institute's research funding comes through grants from agencies in Tokyo. On a project financed by the New Energy Development Organization, scientists at the institute will develop general quantitative models to simulate cell metabolism in *Escherichia coli*. The long-term goal of this five-year project is to develop tools for the engineering and design of microorganisms for industrial applications. Research on the simulation of metabolism in rice is funded by, and undertaken in cooperation with, the rice genome project at the agriculture ministry.

And today, E-CELL is not the isolated project it was a few years ago. "When I started the E-CELL project some six years ago, I would estimate that 99% of bioscientists in Japan did not take my work seriously," recalls Tomita. "Today I would estimate that number at below 50%." As that number declines, the prospects for fresh ways of organizing and running research institutions continue to improve.

Robert Triendl is a freelance writer based in Tokyo.

Institute for Advanced Biosciences ♦ www.ttck.keio.ac.jp/IAB

E-CELL ♦ www.e-cell.org

Virtual world: researchers at the Institute for Advanced Biosciences aim to generate accurate computer models of living cells.